

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 4, 2004, 18:36:21 ; Search time 147.5 Seconds

(without alignments)  
12757.738 Million cell updates/sec

Title: US-10-007-270-1

Perfect score: 5875

Sequence: 1 taacaagaaggttatcct.....tactatatgacataatcaat 3330

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DEV=xl  
-Q/cn2\_1/USPTO.spool\_p/US10007270/runat\_04032004.160715.4884/app\_query.fasta\_1.3527  
-DB=A Geneseg 29Jan04 -QW=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOFCU=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10007270 @CN 1 1 117 @runat\_04032004.160715.4884 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DRV TIMEOUT=120 -WREN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A Geneseg 29Jan04:\*  
1: Genesegp1980s:\*  
2: Genesegp1980s:\*  
3: Genesegp2000s:\*  
4: Genesegp2001s:\*  
5: Genesegp2002s:\*  
6: Genesegp2003s:\*  
7: Genesegp2003bs:\*  
8: Genesegp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score  | Query Match | Length | ID         | Description        |
|------------|--------|-------------|--------|------------|--------------------|
| 1          | 4073   | 69.3        | 797    | 6 ADAL4841 | Adal4841 Human int |
| 2          | 4073   | 69.3        | 797    | 7 ABR42342 | AbR42342 Human int |
| 3          | 3862   | 65.7        | 771    | 3 AAY57089 | Aay57089 Human int |
| 4          | 3862   | 65.7        | 771    | 3 AAY93336 | Aay93336 A human i |
| 5          | 3862   | 65.7        | 771    | 6 ADAL4867 | Adal4867 Human var |
| 6          | 3862   | 65.7        | 771    | 7 ABR42354 | AbR42354 Human int |
| 7          | 3611   | 61.5        | 719    | 6 ADAL4843 | Adal4843 Human int |
| 8          | 3611   | 61.5        | 719    | 7 ABR42343 | AbR42343 Human int |
| 9          | 2504.5 | 42.6        | 798    | 6 ADAL4848 | Adal4848 Mouse int |
| 10         | 2504.5 | 42.6        | 798    | 7 ABR42345 | AbR42345 Mouse int |

|    |        |      |      |            |                    |
|----|--------|------|------|------------|--------------------|
| 11 | 2171   | 37.0 | 709  | 3 AAY93338 | Aay93338 A murine  |
| 12 | 1298.5 | 22.1 | 466  | 6 ADAL4850 | Adal4850 Mouse int |
| 13 | 1298.5 | 22.1 | 466  | 7 ABR42346 | AbR42346 Mouse int |
| 14 | 1038   | 17.7 | 198  | 6 ADAL4845 | Adal4845 Human int |
| 15 | 1038   | 17.7 | 198  | 7 ABR42344 | AbR42344 Human int |
| 16 | 918    | 15.6 | 1241 | 3 AAY93337 | Aay93337 A human i |
| 17 | 917    | 15.6 | 1241 | 6 ADAL4856 | Adal4856 Human int |
| 18 | 917    | 15.6 | 1241 | 7 ABR42349 | AbR42349 Human int |
| 19 | 841    | 14.3 | 185  | 3 AAY93335 | Aay93335 An interp |
| 20 | 841    | 14.3 | 185  | 6 ADAL4854 | Adal4854 Monkey in |
| 21 | 841    | 14.3 | 185  | 7 ABR42348 | AbR42348 Monkey in |
| 22 | 820.5  | 14.0 | 1069 | 3 AAY93339 | Aay93339 A murine  |
| 23 | 820.5  | 14.0 | 1069 | 6 ADAL4863 | Adal4863 Mouse int |
| 24 | 820.5  | 14.0 | 1069 | 7 ABR42352 | AbR42352 Mouse int |
| 25 | 420.5  | 7.2  | 432  | 6 ADAL4858 | Adal4858 Human int |
| 26 | 420.5  | 7.2  | 432  | 7 ABR42350 | AbR42350 Human int |
| 27 | 196    | 3.3  | 577  | 3 AAY71030 | Aay71030 Ubiquitin |
| 28 | 189.5  | 3.2  | 1255 | 5 ABP56040 | Abp56040 MUC1 rece |
| 29 | 189.5  | 3.2  | 1255 | 6 ABB82568 | Abb82568 MUC1/REP  |
| 30 | 189.5  | 3.2  | 1255 | 6 ABR47537 | AbR47537 Breast ca |
| 31 | 189.5  | 3.2  | 1255 | 6 ABR92124 | AbR92124 Human cer |
| 32 | 189.5  | 3.2  | 1255 | 7 ADD45111 | Add45111 Human Pro |
| 33 | 189.5  | 3.2  | 1255 | 7 ADE54622 | Ade54622 Human Pro |
| 34 | 188.5  | 3.2  | 495  | 4 AAU00539 | Aau00539 Human MUC |
| 35 | 188.5  | 3.2  | 515  | 5 ABG96378 | Abg96378 Human ova |
| 36 | 188.5  | 3.2  | 515  | 6 ADA50565 | Ada50565 Human muc |
| 37 | 188.5  | 3.2  | 515  | 6 ABR92123 | AbR92123 Human cer |
| 38 | 188.5  | 3.2  | 515  | 6 AAE37797 | Aae37797 Human muc |
| 39 | 185.5  | 3.2  | 508  | 2 AAU77233 | Aau77233 MiniMUC1  |
| 40 | 183.5  | 3.1  | 455  | 3 AAU71024 | Aay71024 Human MUC |
| 41 | 183.5  | 3.1  | 475  | 4 AAU00573 | Aau00573 Human MUC |
| 42 | 183.5  | 3.1  | 475  | 5 ABE77476 | Abb77476 Human MUC |
| 43 | 183.5  | 3.1  | 475  | 6 ADA50567 | Ada50567 Mucin 1 ( |
| 44 | 183.5  | 3.1  | 475  | 6 AAE37800 | Aae37800 Human muc |
| 45 | 183.5  | 3.1  | 475  | 7 ADD14120 | Add14120 Human src |

## ALIGNMENTS

RESULT 1

ADAL4841

ID ADAL4841 standard; protein; 797 AA.

XX AC ADAL4841;

XX DT 06-NOV-2003 (first entry)

DE Human interphotoreceptor matrix component, IPMC, 150 isoform A.

XX DE human; IPMC 150 isoform A; gene therapy;  
interphotoreceptor matrix component; IPMC; ocular disorder;  
macular degeneration; photoreceptor death; retinal detachment.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide 1..20

FT Protein /note= "Signal sequence"

FT Protein 21..797

FT Misc-difference /note= "Mature IPMC 150 isoform A"

FT Misc-difference 187..188

FT /note= "Encoded by ACACGATCTTCATTTCAAGACTTGGCGAGTATTCTA  
AGAAACCTTCAGAGCAATTCAAGAT"

XX US2002160954-A.

XX 31-OCT-2002.

XX 08-NOV-2001; 2001US-00007270.

XX 29-OCT-1998; 98US-00183972.

XX 29-OCT-1999; 99US-00430195.

XX

PA (IOWA ) UNIV IOWA RES FOUND.  
 XX Hageman GS, Kuehn MH;  
 XX WPI; 2003-238235/23.  
 DR N-PSDB; ADA14840.  
 XX  
 XX New isolated or recombinant interphotoreceptor matrix component  
 PT polynucleotide and polypeptide, useful for diagnosing, preventing,  
 PT treating or prognosticating ocular disorders, e.g. macular degeneration  
 PT or retinal detachment.  
 XX  
 XX Claim 8; Page 31-33; 76pp; English.  
 XX  
 CC The invention relates to an isolated or recombinant interphotoreceptor  
 CC matrix component (IPMC) polynucleotide. Also disclosed is a vector  
 CC comprising a promoter of an interphotoreceptor matrix component (IPMC)  
 CC gene operatively linked to the IPMC polynucleotide. The IPMC  
 CC polynucleotides, polypeptides and antibodies are useful for diagnosing,  
 CC preventing, treating or prognosticating ocular disorders, e.g. macular  
 CC degeneration, photoreceptor death or retinal detachment. They are also  
 CC useful for identifying a compound capable of modulating IPMC gene  
 CC expression in a cell. The present sequence represents the amino acid  
 CC sequence of human interphotoreceptor matrix component, IPMC, 150 isoform  
 CC A.  
 XX  
 SQ Sequence 797 AA;  
 Alignment Scores:  
 Pred. No.: 0 Length: 797  
 Score: 4073.00 Matches: 796  
 Percent Similarity: 97.31% Conservative: 1  
 Best Local Similarity: 97.19% Mismatches: 0  
 Query Match: 69.33% Indels: 22  
 DB: 6 Gaps: 1  
 US-10-007-270-1 (1-3330) x ADA14841 (1-797)  
 QY 131 ATGTATTGGAACTAGAGAGCTATTGTTGTTTGGATTTTCTCAAGTTCAGGA 190  
 Db 1 MetTyrLeuGluThrArgArgAlaIlePheValPheTrpIlePheLeuGlnValGlnGly 20  
 QY 191 ACCAAGATATCTCCATTAAATATACATATACCATCTGAACTAAAGACATAGACATCCCA 250  
 Db 21 ThrTyrAspIleSerIleAenIleTyrHisSerGluThrLysAspIleAspAsnProPro 40  
 QY 251 AGAATGAACAACTGAAGTACTGAAAAATGTCAAAATGTCAACTATCAGACGAATA 310  
 Db 41 ArgAsnGluThrThrGluSerThrGluLysMetTyrLysMetSerThrMetArgArgIle 60  
 QY 311 TTCGATTGGCAAGCATCGAACCAAAAGATCCGATTTTCCCAACGGGGTTAAAGTC 370  
 Db 61 PheAspLeuAlaLysHisArgThrLysArgSerAlaPhePheProThrGlyValLysVal 80  
 QY 371 TGTCACAGGAATCCATGAACAGATTTTACAGAGCTCTTCAAGCTTATTATAGATTGAGA 430  
 Db 81 CysProGlnGluSerMetLysGlnIleLeuAspSerLeuGlnAlaTyrArgLeuArg 100  
 QY 431 GTCTGTTCAGGACGATCGGAAGCATATCGGATCTTCTCGATCGCATCCCTGACACA 490  
 Db 101 ValCysGlnGluAlaValTrpGluAlaTyrArgIlePheLeuAspArgIleProAspThr 120  
 QY 491 GGGGAATATCAGGACTGGGTACAGCATCTGCCAGCAGGACCTTCTGCTTTCGATTT 550  
 Db 121 GlyGluTyrGlnAspTrpValSerIleCysGlnGlnGluThrPheCysLeuPheAspIle 140  
 QY 551 GGAATAATCTCAGCAATCCAGGACACTGGATCTTCTCCAGCAGAGATATAAAGACAG 610  
 Db 141 GlyLysAsnPheSerAsnSerGlnGlnHisLeuAspLeuLeuGlnGlnAlaArgIleLysGln 160  
 QY 611 AGAAGTTTCCCTGACAGAAAGATGAATATCTGCAGAGAGACATTTGGAGAGCGCTGGT 670  
 Db 161 ArgSerPheProAspArgLysAspGluIleSerAlaGluLysThrLeuGlyGluProGly 180

QY 671 GAAACCATTTGTCAATTCACAGCAATCTACATTTCAAAGACTTGGGAGTATTCTTAAGAA 730  
 Db 181 GluThrIleValIleSer--Thr----- 187  
 QY 731 AACCTCAGAGAGCAAAATTCAGAGTGTGGCAACGTCTCACTTGGGCTTTCCTCTCA 790  
 Db 188 -----AspValAlaAsnValSerLeuGlyProPheProLeuT 200  
 QY 791 CTCTGATGACACCTCTCTCAATGAAATTTCTCGATATATACCTCAACACACCAAGATGC 850  
 Db 200 hrProAspAspThrLeuLeuAsnGluIleLeuAspAsnThrLeuAsnAspThrLysMetP 220  
 QY 851 CTACACAGAGAGAGAAACAGAAATTCGCTGTGTGGAGGAGAGAGGTGGAGTCAAGCG 910  
 Db 220 roThrThrGluArgGluThrGluPheAlaValLeuGluGluGluGluGluValGluLeuSerV 240  
 QY 911 TCTCTCTGTAAACACAGAGTTCAAGGACAGAGCTCCGCTGACTCCAGTCCCAATATTACC 970  
 Db 240 alSerLeuValAsnGlnLysPheLysAlaGluLeuAlaAspSerGlnSerProTyrTyrG 260  
 QY 971 AGAGCTAGCAGGAAAGTCCCACTTCAGATGCCAAAGATATTTAAGAAATTCAGAGAT 1030  
 Db 260 lngLiuLeuAlaGlyLysSerGlnLeuGlnMetGlnLysIlePheLysLysLeuProGlyP 280  
 QY 1031 TCAGAAATCCATGTGTAGATTAGACCAAGAAAGAAAGAGTGGCTCAAGCTCCA 1090  
 Db 280 helyLysIleHisValLeuGlyPheArgProLysLysGluLysAspGlySerSerSert 300  
 QY 1091 CAGAGATCAACTTACGGCATCTTTAAGAGACAGAGTGCAGAGACCAAGAGCCCTGCAA 1150  
 Db 300 hrGluMetGlnLeuThrAlaIlePheLysArgHisSerAlaGluAlaLysSerProAla 320  
 QY 1151 GTGACCTCTGTCTTTGATTCCCAAAATTTGAAGTGCAGAAAGTCTATCATGCAACCA 1210  
 Db 320 erAspLeuLeuSerPheAspSerAenLysIleGluSerGluGluValTyrHisGlyThrM 340  
 QY 1211 TCGAGAGGACAGAGCAACCAAGATCTATCTCAGAGTACAGACCTCAAAAGCTGATCA 1270  
 Db 340 eLysGluAspLysGlnProGluIleTyrLeuThrAlaThrAspLeuLysArgLeuIleS 360  
 QY 1271 GCAAGACATAGAGAGAAACAATCTTTGATGTGGGTCGGGCAATTCAGTTCATGATGAAA 1330  
 Db 360 erLysAlaLeuGluGluGlnSerLeuAspValGlyThrIleGlnPheThrAspGluI 380  
 QY 1331 TTGCTGATCATCGCAGCCCTTGGTCTGCACACCAATCAGAGCTGCCCACTTTTTCG 1390  
 Db 380 leAlaGlySerLeuProAlaPheGlyProAspThrGlnSerGluLeuProThrSerPheA 400  
 QY 1391 CTGTTTATACAGAGATGCTACTTTCAGTCCAGAGACTTCTCTGTGAAACCCAGCTTG 1450  
 Db 400 laValIleThrGluAspAlaThrLeuSerProGluLeuProValGluProGlnLeuG 420  
 QY 1451 AGACAGTGCAGGACAGAGATGCTTACCTGACACTTCTGCTCCACCTCCCTATGCG 1510  
 Db 420 luthrValAspGlyAlaGluHisGlyLeuProAspThrSerTrpSerProAlaMetA 440  
 QY 1511 CCTTACCTTCCCTGTTCAGAGCTCCACCTTTCTTATGTCATCAAGCATCTTCTCTCTGA 1570  
 Db 440 laSerThrSerLeuSerGluAlaProProPhePheMetAlaSerSerIlePheSerLeuT 460  
 QY 1571 CTGATCAAGGACACACAGATACATGGCCACTGACAGACCAATGCTAGTACAGGCTCA 1630  
 Db 460 hrAspGlnGlyThrThrAspThrMetAlaThrAspGlnThrMetLeuValProGlyLeuT 480  
 QY 1631 CCATCCCCACAGTGAATTTCTGCAATCAGCCAACTGGCTCTGGGAATTTTCATATCCAC 1690  
 Db 480 hrIleProThrSerAspTyrSerAlaIleSerGlnLeuAlaLeuGlyIleSerHisProP 500  
 QY 1691 CTGCATCTTCAGATCAGCCCATCAAGTGCAGGTGGGAGATATGTCAGACACCTAG 1750  
 Db 500 roAlaSerSerAspAspSerArgSerSerAlaGlyGluAspMetValArgHisLeuA 520

QY 1751 ATGAATGATCTGTCTGACACATCTCTGCCCCATCTGAGTACCGAGCTCAGGCAATATG 1810  
 Db |||||  
 520 spGluMetAspLeuSerAspThrProAlaProSerGluValProGluLeuSerGluTyrV 540  
 QY 1811 TTTCTGTCCAGATCATCTTCTGGAGATACCACTCTCTCTCTCAGCTTACAGTATATCA 1870  
 Db |||||  
 540 alSerValProAspHisPheLeuGluAspThrThrProValSerAlaLeuGluTyrIleT 560  
 QY 1871 CCATAGTCTTATGACCATTCGCCAGGCGGCGAGAGCTGTAGTCTTCTCAGTCTGC 1930  
 Db |||||  
 560 hrThrSerSerMetThrIleAlaProLysGlyArgGluLeuValValPhePheSerLeuA 580  
 QY 1931 GTGTGTCTTAACATGGCTCTTCCACAGCACTGTTCAACAGAGCTCTCTGAGTACCCAG 1990  
 Db |||||  
 580 rgValAlaAsnMetAlaPheSerAsnAspLeuPheAsnLysSerSerLeuGluTyrArg 600  
 QY 1991 CTTCTGGAGCAACATTCACACAGCTGTCTTCCATCTACGATCCATCTTACAGAT 2050  
 Db |||||  
 600 laLeuGluGlnGlnPheThrGlnLeuLeuValProTyrLeuArgSerAsnLeuThrGlyP 620  
 QY 2051 TTAAGCAACTGAAATACCTTAACCTCAGAACGGAGTGTGATTTGATGATACAAATGA 2110  
 Db |||||  
 620 heLysGlnLeuGluLeuLeuAsnPheArgAsnGlySerValIleValAsnSerLysMetL 640  
 QY 2111 AGTTTGTAGTCTGCGGTATTAACCTCAGAACGGTGTGCGAGGGCTTGGAGGATT 2170  
 Db |||||  
 640 ysPheAlaLysSerValProTyrAsnLeuThrLysAlaValHisGlyValLeuGluAsp 660  
 QY 2171 TTTCTGTCTGCGAGCCCAACACTCTGGAATAGACAGTCTCTCTCAACATG 2230  
 Db |||||  
 660 heArgSerAlaAlaGlnGlnLeuHisLeuGluLeuAspSeryrSerLeuAsnIleG 680  
 QY 2231 AACAGCTGATCAAGCAGATCTCTGCAAGTCTCTGCGCTGCGCGAATTCGCCAATGTG 2290  
 Db |||||  
 680 luProAlaAspGlnAlaAspProCysPheLeuAlaCysGlyGluPheAlaGlnCysV 700  
 QY 2291 TAAAGAACGAGCTGAGAGAGCGAGTGTCTGCTCAACAGAGATATCACAGCCAGG 2350  
 Db |||||  
 700 alLysAsnGluArgThrGluGluAlaGluCysArgCysLysProGlyrAspSerGlnG 720  
 QY 2351 GGAGCTGTGAGCGTGTGACACAGCGCTCTGTGCTGCGCTGCGCAAGAGTGGAGTCC 2410  
 Db |||||  
 720 lySerLeuAspGlyLeuGluProGlyLeuCysGlyProGlyThrLysGluCysGluVal 740  
 QY 2411 TCCAGGGAAGGAGCTCCATGCGAGTGTGCGAGATCACTCTGAAATCAAGCATACAAA 2470  
 Db |||||  
 740 euGlnGlyLysGlyAlaProCysArgLeuProAspHisSerGluAsnGlnAlaTyrLysT 760  
 QY 2471 CTAGTGTAAAAGTTCCAAAATCAACAAATACAGGTAATCAAGTAAAGAAATCTGT 2530  
 Db |||||  
 760 hrSerValLysPheGlnAsnGlnAsnLysValIleSerLysArgAsnSerG 780  
 QY 2531 AATTACTGACGTAGATATGAGATTTAACATCAAGATTTGGAGGAAT 2583  
 Db |||||  
 780 luLeuLeuThrValGluTyrGluGluPheAsnHisGlnAspTrpGluGlyAsn 797

RESULT 2

ID ABR42342 standard; protein; 797 AA.  
 AC ABR42342;  
 XX

DT 11-AUG-2003 (first entry)

DE Human interphotoreceptor matrix IPM 150, isoform A.

KW Human; interphotoreceptor matrix; IPM 150; IPMC; receptor;  
 KW opthalmological; gene therapy.

OS Homo sapiens.

FH Key Location/Qualifiers  
 FT Peptide 1. .20

Protein /label= Signal\_peptide  
 FT 21..797  
 Domain /label= IPM 150  
 FT 71..88  
 Domain /note= "N-terminal domain"  
 FT 95..115  
 Domain /note= "conserved domain"  
 FT 187..188  
 Misc-difference /note= "encoded by ACAGCAATCTACATTTCCAAAAGACTTGGCAG  
 TATCTTAAGAAACCTCAGAGGACAAATTCAGAT"  
 FT 688..731  
 Domain /note= "EGF-like domain"  
 FT XX  
 PN WO2003039346-A2.  
 XX  
 PD 15-MAY-2003.  
 XX  
 PF 08-NOV-2002; 2002WO-US036090.  
 XX  
 PR 08-NOV-2001; 2001US-00077270.  
 XX  
 PA (IOWA ) UNIV IOWA RES FOUND.  
 XX  
 PI Hageman GS, Kuehn MH;  
 XX  
 DR WPI; 2003-441440/41.  
 DR N-PSDB; ACC57946.  
 XX  
 XX New interphotoreceptor matrix proteins and polynucleotides, useful for  
 treating or preventing photoreceptor death or retinal detachment, or for  
 treating ocular disorders.  
 Claim 6; Page 78; 105pp; English.

The present sequence is the protein sequence of isoform A of novel human

interphotoreceptor matrix 150 (IPM 150), a member of the newly identified  
 interphotoreceptor matrix component (IPMC) gene family. The IPM 150 gene  
 is located on chromosome 6q13-q15, a region that also contains loci for  
 progressive bifocal choriorretinal atrophy, autosomal dominant Stargardt's  
 -like macular dystrophy, North Carolina macular dystrophy and Salla  
 disease. Members of the IPMC gene family have been identified in humans,  
 monkey, cow, goat, rabbit, dog, cat, pig, mouse and rat IPM. 2  
 Subfamilies are designated IPM 150 (or IPMGI) and IPM 200 (or IPMG2). The  
 invention provides IPM 150 and IPM 200 polynucleotides and polypeptides,  
 antibodies that specifically bind the polypeptides, and vectors  
 comprising the polynucleotides. A claimed method of treating or  
 preventing photoreceptor death or retinal detachment involves  
 administering an IPMC polynucleotide, polypeptide or antibody. Also  
 claimed is a method for identifying a compound capable of modulating IPMC  
 gene expression

SQ Sequence 797 AA;

Alignment Scores:

Pred. No.: 0 Length: 797  
 Score: 4073.00 Matches: 796  
 Percent Similarity: 97.31% Conservative: 1  
 Best Local Similarity: 97.19% Mismatches: 0  
 Query Match: 69.33% Indels: 22  
 DB: 7 Gaps: 1

US-10-007-270-1 (1-3330) x ABR42342 (1-797)

QY 131 ATGTATTGGAACTAGAGAGCTATTTGTTTTGGATTTTCTCCAAAGTCAAGA 190  
 Db |||||  
 1 MetTyrLeuGluThrArgAlaIlePheValPheTrpIlePheLeuGlnValGlnGly 20  
 QY 191 ACCAAAGATATCTCCATTAACATATACCATCTCTGAACCTAAGACATAGCAATCCCCCA 250  
 Db |||||  
 21 ThrLysAspIleSerIleAsnIleTyrHisSerGluThrLysAspIleAsnProPro 40  
 QY 251 AGAAATGAAACAATCGAAAGTACTGAAATAATGTCAAAATGTCAACTATGAGAGATA 310  
 Db |||||

Db 41 ArgAsnGluThrThrGluSerThrGluLysMetTyrLysMetSerThrMetArgArgIle 60  
QY 311 TTCCGATTGGCAAGACATCGAACAAAGATCCGCAATTTTCCCAACGGGGTAAAGTC 370  
Db 61 PheAspLeuAlaLysHisArgThrLysArgSerAlaPhePheProThrGlyValIleVal 80  
QY 371 TGTCCACAGGAATCCATGAACAGATTTTACAGATCTTCAAGCTTATATAGATTGAGA 430  
Db 81 CysProGlnGluSerMetLysGlnIleLeuAspSerLeuGlnAlaTyrTyrArgLeuArg 100  
QY 431 GTGTCTCAGGAGCAGTATGGAGCATATCGGATCTTCTGGATCGCATCCCTGCACACA 490  
Db 101 ValCysGlnGluAlaValTrpGluAlaTyrArgIlePheLeuAspArgIleProAspThr 120  
QY 491 GGGGAATATCAGGACTGGCTCAGCATCTGCCAGCAGGAGACTTCTGCTCTTTGCATT 550  
Db 121 GlyGluTyrGlnAspTrpValSerIleCysGlnGlnGluThrPheCysLeuPheAspIle 140  
QY 551 SGAAAAACTTCAGCAATTCGAGGACCTGGATCTTCTCCACAGACAGATAAAGACAG 610  
Db 141 GlyLysAsnPheSerAsnSerGlnGluHisLeuAspLeuLeuGlnArgIleLysGln 160  
QY 611 AGAAGTTTCCCTCAGCAGAAAGATGAATATCTGCAGAGAGACATTTGGAGAGCTGGT 670  
Db 161 ArgSerPheProAspArgLysAspGluIleSerAlaGluLysThrLeuGlyGluProGly 180  
QY 671 GAAACCATTTCTCAATTTCAACAGCAATCTACATTTCAAGACTTGGCGCATTTCTAAGAA 730  
Db 181 GluThrIleValIleSer--Thr----- 187  
QY 731 AACCTCAGAGAGCAAAATTCAGATGTTGCCAAGCTCTCACTTGGGCTTTCCCTCTCA 790  
Db 188 -----AspValAlaAsnValSerLeuGlyProPheProLeu 200  
QY 791 CTCCTGATGACACCTCTCAATGAATTCGATATACATCTCAACGACACCAAGATGC 850  
Db 200 hrProAspThrLeuLeuAsnGluIleLeuAspAsnThrLeuAsnAspThrLysMetP 220  
QY 851 CTACACAGAAAGAAACAGAAATTCGCTGTGTTGGAGGAGCAGAGGCTGAGCTCAGG 910  
Db 220 rothrThrGluArgGluThrGluPheAlaValLeuGluGluGluArgValGluLeuSerV 240  
QY 911 TCTCTCTGTAAACAGAGTTCAGGAGCTGCTGACTCCGACCTCCGACCTCCCATATTACC 970  
Db 240 alSerLeuValAsnGlnLysPheLysAlaGluLeuAlaAspSerGlnSerProTyrTyrG 260  
QY 971 AGGAGTACGAGAAAGTCCCACTTCAGATGCAAGATATTTAAGAAATTTCCAGGAT 1030  
Db 260 lngluLeuAlaGlyLysSerGlnLeuGlnMetGlnLysIlePheLysLysLeuProGlyP 280  
QY 1031 TCMAAAATCCATGTGTAGATTAGACCAAGAAAGAAAGAAAGATGCTCAAGCTCCA 1090  
Db 280 helysLysIleHisValLeuGlyPheArgProLysLysGluLysAspGlySerSerSert 300  
QY 1091 CAGAGATCACTTACGGCCATCTTAAGAGACACAGTGCAGAGCAAAAGCCCTGCAG 1150  
Db 300 hrGluMetGlnLeuThrAlaIlePheLysArgHisSerAlaGluAlaLysSerProAlaS 320  
QY 1151 GTGACCTCTGCTTTTGTATCCCAAAATTCGAAGTGAAGAGTCTATCATGAACCA 1210  
Db 320 erAspLeuLeuSerPheAspSerAsnLysIleGluSerGluGluValTyrHisGlyThrM 340  
QY 1211 TGGAGGAGCAGAGCAACAGAAATCTATCTCAGCTACAGACTCAAGCCCTCAAAAGGCTGATCA 1270  
Db 340 etGluGluAspLysGlnProGluIleTyrLeuThrAlaThrAspLeuLysArgLeuIleS 360  
QY 1271 GCAAGCACTAGAGGAGAGAACATCTTTGGATGTGGACAAATTCAGTTCACTGATGAAA 1330  
Db 360 erLysAlaLeuGluGluGlnSerLeuAspValGlyThrIleGlnPheThrAspGluI 380  
QY 1331 TTGCTGGATCACTGCCAGCTTTGCTCTCAGACCCATCAGAGCTGCCCACTTTTG 1390  
Db 380 leAlaGlySerLeuProAlaPheGlyProAspThrGlnSerGluLeuProThrSerPheA 400

QY 1391 CTGTTTATACAGAGGATGCTACTTTGAGTCCAGAACTTCTCTCTGTGAACCCAGCTTG 1450  
Db 400 laValIleThrGluAspAlaThrLeuSerProGluLeuProValGluProGlnLeuG 420  
QY 1451 AGACAGTGCAGGAGCAGACATGTCTACTCTGACACATCTTTGGTCTCCACTGCTATGG 1510  
Db 420 luthrValAspGlyAlaGluHisGlyLeuProAspThrSerTrpSerProAlaMetA 440  
QY 1511 CTTTACTCTCTCTCAGAAAGCTCCACCTTTCTTTATGCAATCAAGCATCTCTCTCTCA 1570  
Db 440 laSerThrSerLeuSerGluAlaProProPhePheMetAlaSerSerIlePheSerLeu 460  
QY 1571 CTGATCAGGACACACAGATACAAATGGCCACTGACAGACAATGCTAGTACACAGGCTCA 1630  
Db 460 hrAspGlnGlyThrAspThrMetAlaThrAspGlnThrMetLeuValProGlyLeu 480  
QY 1631 CCATCCCCACAGTGTATTCTGCAATCAGCAACCTGGCTCTGGGAATTTACATCCAC 1690  
Db 480 hrIleProThrSerAspTyrSerAlaIleSerGlnLeuAlaLeuGlyIleSerHisProp 500  
QY 1691 CTGCATCTTCAGATGACAGCCGATCAAGTGCAGGTGGGAAAGATATGCTCAGACACCTAG 1750  
Db 500 roAlaSerSerAspAspSerArgSerAlaGlyGlyGluAspMetValArgHisLeuA 520  
QY 1751 ATGAAATGGATCTGCTGACACATCTCTGCCCATCTGAGGTACAGAGCTCAGCGAATATG 1810  
Db 520 spGluMetAspLeuSerAspThrProAlaProSerGluValProGluLeuSerGluTyrV 540  
QY 1811 TTTCTGTCCAGATCAATTTCTTGAGGATACCACTCTCTGCTCTCAGCTTTACAGTATATCA 1870  
Db 540 alSerValProAspHisPheLeuGluAspThrThrProValSerAlaLeuGlnTyrIle 560  
QY 1871 CCACTACTTCTATGACCATTCGCCCAAGGCGGAGCTGGTAGTGTCTTCTCACTGCTGC 1930  
Db 560 hrThrSerSerMetThrIleAlaProLysGlyArgGluLeuValValPhePheSerLeuA 580  
QY 1931 GTGTGTCTAACATGCGCTTCTCCACAGCTTTTCAACAGAGCTCTCTGGAGTACCGAG 1990  
Db 580 rgValAlaAsnMetAlaPheSerAsnAspLeuPheAsnLysSerSerLeuGluTyrArg 600  
QY 1991 CTCTGGAGCAACAATTCACACAGCTCTGCTGTTCATATCTAGCATCCAATCTTACAGGT 2050  
Db 600 laLeuGluGlnPheThrGlnLeuLeuValProTyrLeuArgSerAsnLeuThrGlyP 620  
QY 2051 TTAAGCACTTCAATATCTTAACCTCAGAAACGGAGTGTGATTTGTAATGACAAATCA 2110  
Db 620 helysGlnLeuGluIleLeuAsnPheArgAsnGlySerValIleValAsnSerLysMetL 640  
QY 2111 AGTTTCTAAGTCTGTCCCTATACCTCACCAGGCTGTGCACGGGTCTTGGAGATT 2170  
Db 640 ysPheAlaLysSerValProTyrAsnLeuThrLysAlaValHisGlyValLeuGluAsp 660  
QY 2171 TTGTTCTGTGCGAGCCCAACATCTCCTGGAATAGACAGCTACTCTCTCAACATG 2230  
Db 660 heArgSerAlaAlaGlnGlnLeuHisLeuGluIleAspSerTyrSerLeuAsnIleG 680  
QY 2231 AACCACTGATCAGCAGATCTCTGCAAGTCTTGGCTTGGCGGCAATTTGCCAATGTG 2290  
Db 680 luProAlaAspGlnAlaAspProCysLysPheLeuAlaCysGlyGluPheAlaGlnCysV 700  
QY 2291 TAAAGACAGCAAGCTGAGGAGCGGAGTGTGCTGCAACACAGGATATGACAGCCAGG 2350  
Db 700 alLysAsnGluArgThrGluGluAlaGluCysArgCysLysProGlyTyrAspSerGlnG 720  
QY 2351 GGAGCTGGAGCGTCTGGAAACAGGCTCTGTGGCCCTGGGCACAAAGGATCCGAGGTCC 2410  
Db 720 lysLeuAspGlyLeuGluProGlyLeuCysGlyProGlyThrLysGluCysGluValI 740  
QY 2411 TCCAGGGAAGGAGCTCCATCGAGGTTCGCGAGATCACTCTGAAAATCAAGATACAAA 2470  
Db 740 euGlnGlyLysGlyAlaProCysArgLeuProAspHisSerGluAsnGlnAlaTyrLysT 760



|                        |  |   |          |          |             |                 |         |      |
|------------------------|--|---|----------|----------|-------------|-----------------|---------|------|
| QY                     | 2471   | CTAGTGTAAAGTTC  | CAAAATCA | CAAAATTA | CAAGGTAT    | CAGTAA          | GAATTCG | 2530 |
| Db                     | 760  | hrSerVallyslysrPheGlnAsnGlnGlnAsnAsnLysValIleSerIlyArgAsnSerG | 780      |          |             |                 |         |      |
| QY                     | 2531   | AATTA   | CTGACCGT | AGATAT   | GAAGAATTTAA | CCATCAAGATTGCGA | AGGAAT  | 2583 |
| Db                     | 780  | luLeulleuthrValGluIytrGluGluPheAsnHisGlnAspTrpGluGlyAsn       | 797      |          |             |                 |         |      |
| RESULT 3               |  |   |          |          |             |                 |         |      |
| ID                     | AAAY57089  |   |          |          |             |                 |         |      |
| XX                     | AAAY57089  | standard; protein; 771 AA.                                    |          |          |             |                 |         |      |
| XX                     | AAAY57089;   |   |          |          |             |                 |         |      |
| DT                     | 28-FEB-2000  | (first entry)   |          |          |             |                 |         |      |
| DE                     | Human interphotoreceptor matrix proteoglycan 1 amino acid sequence.              |   |          |          |             |                 |         |      |
| XX                     | Neglected target tissue antigen; NNTA; autoimmunity; autoimmune response;        |   |          |          |             |                 |         |      |
| XX                     | immunotherapeutic agent; insulin dependent diabetes mellitus;                    |   |          |          |             |                 |         |      |
| XX                     | multiple sclerosis; autoimmune thyroiditis; rheumatoid arthritis;                |   |          |          |             |                 |         |      |
| XX                     | uveoretinitis; inflammatory response.  |   |          |          |             |                 |         |      |
| OS                     | Homo sapiens.  |   |          |          |             |                 |         |      |
| XX                     | WO9956763-A1.  |   |          |          |             |                 |         |      |
| XX                     | 11-NOV-1999.   |   |          |          |             |                 |         |      |
| XX                     | 07-MAY-1999;   | 99WO-US010250.  |          |          |             |                 |         |      |
| XX                     | 07-MAY-1998;   | 98US-0084636P.  |          |          |             |                 |         |      |
| XX                     | (REGC ) UNIV CALIFORNIA.   |   |          |          |             |                 |         |      |
| XX                     | Kaufman DL, Tian J, Olcott A;  |   |          |          |             |                 |         |      |
| XX                     | WPI; 2000-052905/04.   |   |          |          |             |                 |         |      |
| XX                     | Administration of neglected target tissue antigens to modulate immune responses. |   |          |          |             |                 |         |      |
| XX                     | Disclosure; Page 30; 79pp; English.  |   |          |          |             |                 |         |      |
| XX                     | Amino acid sequences AAAY57063-Y57091 are examples of neglected target           |   |          |          |             |                 |         |      |
| XX                     | tissue antigens NNTAs. NNTAs are antigens (whole antigens or fragments)          |   |          |          |             |                 |         |      |
| XX                     | not involved in autoimmunity. These peptides and proteins are used in the        |   |          |          |             |                 |         |      |
| XX                     | method of the invention which involves administering an NNTA as an               |   |          |          |             |                 |         |      |
| XX                     | antigen based immunotherapeutic agent, to a host afflicted with an               |   |          |          |             |                 |         |      |
| XX                     | autoimmune response associated with an autoimmune disease. The                   |   |          |          |             |                 |         |      |
| XX                     | immunotherapeutic agent is used to treat autoimmune diseases such as             |   |          |          |             |                 |         |      |
| XX                     | insulin dependent diabetes mellitus, multiple sclerosis, autoimmune              |   |          |          |             |                 |         |      |
| XX                     | thyroiditis, uveoretinitis, rheumatoid arthritis or abnormal inflammatory        |   |          |          |             |                 |         |      |
| XX                     | immune responses. The NNTA induces regulatory tolerance by elicitation of        |   |          |          |             |                 |         |      |
| XX                     | regulatory T cells among T cells recognizing the NNTA but not                    |   |          |          |             |                 |         |      |
| XX                     | participating in the immune response. The NNTA are capable of recognition        |   |          |          |             |                 |         |      |
| XX                     | by substantial populations of uncommitted T cells which can be primed, or        |   |          |          |             |                 |         |      |
| XX                     | biased, towards regulatory responses to provide effective treatment. The         |   |          |          |             |                 |         |      |
| XX                     | NNTA are effective in regulating undesirable immune responses even when          |   |          |          |             |                 |         |      |
| XX                     | target determinants used as agents promoting tolerance agents have failed        |   |          |          |             |                 |         |      |
| XX                     | to induce an effective regulatory T cell response. NNTAs as agents               |   |          |          |             |                 |         |      |
| XX                     | promoting tolerance are anticipated to be safer than use of target               |   |          |          |             |                 |         |      |
| XX                     | determinants   |   |          |          |             |                 |         |      |
| XX                     | Sequence 771 AA;   |   |          |          |             |                 |         |      |
| XX                     | SQ   |   |          |          |             |                 |         |      |
| Alignment Scores:      |  |   |          |          |             |                 |         |      |
| Pred. No.:             | 0  | Length:   | 771      |          |             |                 |         |      |
| Score:                 | 3862.00  | Matches:  | 769      |          |             |                 |         |      |
| Percent Similarity:    | 96.74%   | Conservative:   | 2        |          |             |                 |         |      |
| Best Local Similarity: | 96.49%   | Mismatches:   | 0        |          |             |                 |         |      |
| Query Match:           | 65.74%   | Gaps:   | 26       |          |             |                 |         |      |
| DB:                    | 3  | Indels:   | 1        |          |             |                 |         |      |

Db 320 erAspLeuLeuSerPheAsnLysIleGluSerGluGluValTyrHisGlyThrM 340  
QY 1211 TGGAGGAGGACAGCAACAGAAATCTATCTACAGCTACAGACCTCAAAAGGCTCATCA 1270  
Db 340 etGluGluAspGlnProGluIleTyrLeuThrAlaThrAspLeuLysArgLeuIleS 360  
QY 1271 GCAAGCACTAGAGAGAGAACATCTTTGGATCTGGGACAAATTCAGTTCCTGATGAAA 1330  
Db 360 erLysAlaLeuGluGluGlnSerLeuAspValGlyThrIleGlnPheThrAspGluI 380  
QY 1331 TTCTGGATCTCTCCAGGCTTTGGTCTGACACCCCAATCAGAGCTCCACATCTTTTG 1390  
Db 380 leAlaGlySerLeuProAlaPheGlyProAspThrGlnSerGluLeuProThrSerPheA 400  
QY 1391 CTGTTTATACAGAGGATCTACTTTAGTCCAGATCTCTCTCTGTTGAACCCAGCTTG 1450  
Db 400 laValIleThrGluAspAlaThrLeuSerProGluLeuProValGluProGlnLeuG 420  
QY 1451 AGACAGTGGAGGAGCAGAGCATGTCTACCTGACACTTCTGTCTCCACCTGCTATGG 1510  
Db 420 luThrValAspGlyAlaGluHisGlyLeuProAspThrSerTrpSerProAlaMetA 440  
QY 1511 CCTCTACTCTCTGTGACAGCTCCACCTTTCTTTATGGCATCAAGCATCTTCTCTGA 1570  
Db 440 laSerThrSerLeuSerGluAlaProPhePheMetAlaSerSerIlePheSerLeuT 460  
QY 1571 CTGATCAAGCACCCACAGATACATGCCACTGACAGACATGCTAGTACAGGGCTCA 1630  
Db 460 hrAspGlnGlyThrThrAspThrMetAlaThrAspGlnThrMetLeuValProGlyLeuT 480  
QY 1631 CCATCCCAACAGATATTCTGCAATCAGCCAACTGGCTCTGGCAATTCACATCCAC 1690  
Db 480 hrIleProThrSerAspTyrSerAlaIleSerGlnLeuAlaLeuGlyIleSerHisProP 500  
QY 1691 CTGCACTTCAGATCAGACCGATCAAGTGCAGTGGCGAAGATATGTTCAGACACTAG 1750  
Db 500 roAlaSerSerAspAspSerArgSerSerAlaGlyGlyGluAspMetValArgHisLeuA 520  
QY 1751 ATGAAATGATCTGTCTGACACTCTGCCCTGCTGAGTACAGAGCTCTCGAGTACCGAG 1810  
Db 520 spGluMetAspLeuSerAspThrProAlaProSerGluValProGluLeuSerGluTyrV 540  
QY 1811 TTCTGTCCAGATCAATTTCTGGAGATACCACTCTCTGCTCAGCTGTGTCTTCTCAGTCTGC 1930  
Db 560 hrThrSerSerMetThrIleAlaProLysGlyArgGluLeuValPhePheSerLeuA 580  
QY 1931 GTGTTGCTAAATGCGCTCTCCACGACCTGTTCACAGAGCTCTCTGGAGTACCGAG 1990  
Db 580 rgValAlaAsnMetAlaPheSerAsnAspLeuPheAsnLysSerSerLeuGluTyrArgA 600  
QY 1991 CTCTGGACCAATTCACAGCTGTCTGTTCCATATCTAGATCCAACTCTTACAGAT 2050  
Db 600 laLeuGluGlnGlnPheThrGlnLeuLeuValProTyrLeuArgSerAsnLeuThrGlyP 620  
QY 2051 TTAAGCAACTTGAATCTACTTACAGAAACGGAGTGTGTGATTTGCAATAGCAAAATGA 2110  
Db 620 heLysGlnLeuGluIleLeuAsnPheArgAsnGlySerValIleValAsnSerLysMetL 640  
QY 2111 AGTTTGTCTAGTCTGTGCGGTATACTCAACCTCAACAGGCTGTGCAGGGGCTTGGAGATT 2170  
Db 640 yaPheAlaLysSerValProTyrAsnLeuThrLysAlaValHisGlyValLeuGluAspP 660  
QY 2171 TTCGTTCTGTGAGCCCAACCACTCCATCTGGAATAGACAGCTACTCTCTCAACATTC 2230  
Db 660 heArgSerAlaAlaAlaGlnGlnLeuHisLeuGluIleAspSerTyrSerLeuAsnIleG 680  
QY 2231 AACAGCTGATCAGCAGATCCCTGCAAGTCTCTGGCTGGCGGAATTTGCCAATGTG 2290  
Db 680 luProAlaAspGlnAlaAspProCysLysPheLeuAlaCysGlyGluPheAlaGlnCysV 700

QY 2291 TAAAGAACCAACGAGTGGAGAGCGAGTCTGCTCAAAACCCAGCATATGACAGCCAGG 2350  
Db 700 alLysAsnGluArgThrGluGluAlaGluCysArgCysLysProGlyTyrAspSerGlnG 720  
QY 2351 GGAGCTGACAGGCTCTGACACCGAGCTCTGTGGCCCTGGCACAAGGAATGCGAGGTCC 2410  
Db 720 lySerLeuAspGlyLeuGluProGlyLeuCysGly-LeuAlaGlnArgAsnAlaArgSer 739  
QY 2411 TCCAGGAAAGGAGCTCCATGTCAGTTCGCCAGATCACTCTGAAATCAGCATACAAA 2470  
Db 740 SerArgGluArgGluLeuHisAla-ValProAspHisSerGluAsnGlnAlaTyrLysT 759  
QY 2471 CTAGTCTTAAAGATTCACAAATCAACAAATACCAAGG 2509  
Db 759 hrSerVal-LysSerSerLysIleAsnLysIleThrArg 771  
RESULT 4  
AAV93336  
ID AAV93336 standard; protein; 771 AA.  
XX  
AC AAV93336;  
XX  
DT 04-SEP-2000 (first entry)  
XX  
DE A human interphotoreceptor matrix proteoglycan (IPM150).  
KW Interphotoreceptor matrix; IPM; proteoglycan; IPM150; IPMC; IPM200;  
KW chromosome 6q13-q15; ocular disease; retinal detachment;  
KW chloroquine retinopathy; retinal degeneration; cone degeneration;  
KW age related macular degeneration; photoreceptor degeneration;  
KW retinal pigment epithelium degeneration; mucopolysaccharidosis;  
KW rod- cone dystrophy; cone-rod dystrophy.  
XX Homo sapiens.  
OS  
FH  
Key Location/Qualifiers  
Domain 42..215  
/note= "domain A"  
FT Modified-site 42  
/note= "potential glycosylation site"  
FT Modified-site 143  
/note= "potential glycosylation site"  
FT Modified-site 191  
/note= "potential glycosylation site"  
FT Modified-site 215  
/note= "potential glycosylation site"  
FT Domain 221..565  
/note= "domain B"  
FT Modified-site 221  
/note= "potential O-linked glycosylation site"  
FT Modified-site 222  
/note= "potential O-linked glycosylation site"  
FT Modified-site 406  
/note= "potential O-linked glycosylation site"  
FT Modified-site 433  
/note= "potential O-linked glycosylation site"  
FT Modified-site 435  
/note= "potential O-linked glycosylation site"  
FT Modified-site 441  
/note= "potential O-linked glycosylation site"  
FT Modified-site 442  
/note= "potential O-linked glycosylation site"  
FT Modified-site 443  
/note= "potential O-linked glycosylation site"  
FT Modified-site 445  
/note= "potential O-linked glycosylation site"  
FT Modified-site 470  
/note= "potential O-linked glycosylation site"  
FT Modified-site 497  
/note= "potential O-linked glycosylation site"  
FT Modified-site 527  
/note= "potential O-linked glycosylation site"

FT Modified-site 529 /note= "potential O-linked glycosylation site"  
 FT Modified-site 537 /note= "potential O-linked glycosylation site"  
 FT Modified-site 561 /note= "potential O-linked glycosylation site"  
 FT Modified-site 565 /note= "potential O-linked glycosylation site"  
 FT Domain 591..630 /note= "domain C"  
 FT Modified-site 591 /note= "potential N-linked glycosylation site"  
 FT Modified-site 592 /note= "potential N-linked glycosylation site"  
 FT Modified-site 616 /note= "potential N-linked glycosylation site"  
 FT Modified-site 630 /note= "potential N-linked glycosylation site"  
 FT Domain 688..731 /note= "EGF-like domain"  
 FT Misc-difference 762 /note= "encoded by AAAA"  
 XX  
 PN WO200026367-A2.  
 XX  
 XX 11-MAY-2000.  
 XX  
 XX 29-OCT-1999; 99WO-US025440.  
 XX  
 XX 29-OCT-1998; 98US-00183972.  
 XX  
 PA (IOWA ) UNIV IOWA RES FOUND.  
 XX  
 PI Hageman GS, Kuehn MH;  
 DR WPI: 2000-365616/31.  
 DR N-PSDB; AAA46205.  
 XX  
 PT Nucleic acids encoding interphotoreceptor matrix proteoglycans useful for preventing, diagnosing and treating ocular disorders such as retinal detachment and choriorretinal degeneration.  
 PT  
 XX  
 PS Claim 5; Fig 3; 183pp; English.  
 XX  
 CC The present sequence represents an interphotoreceptor matrix (IPM) proteoglycan, designated IPM150. The protein is an IPM component (IPMC). Two subfamilies of IPMCs, IPM150 and IPM200, exist. The human IPM150 gene is located on chromosome 6q13-q15, between markers CHLC.GATALLF10 and D6S284. The IPM proteins may be used to supplement a patient's own production of the protein or to rectify alterations in their nucleic acids that result in expression of an inactive protein. The IPM nucleic acids may be used in this way to treat ocular diseases such as retinal detachment, choriorretinal degeneration, RPE (retinal pigment epithelium) degeneration, photoreceptor degeneration, age related macular degeneration, cone degeneration, mucopolysaccharidosis, rod-cone dystrophy and cone-rod dystrophy. The nucleic acids and proteins may also be used to assay for other modulators of IPM proteoglycan expression and activity that may be used to treat ocular diseases. The nucleic acids and proteins may also be used as diagnostic reagents to detect the presence of IPM nucleic acids and their products in samples from patients according to standard methodologies  
 XX  
 SQ Sequence 771 AA;  
 Alignment Scores:  
 Pred. No.: 0 Length: 771  
 Score: 3862.00 Matches: 769  
 Percent Similarity: 96.74% Conservative: 2  
 Best Local Similarity: 96.49% Mismatches: 0  
 Query Match: 65.74% Indels: 26  
 DB: 3 Gaps: 1

US-10-007-270-1 (1-3330) x AAY93336 (1-771)  
 QY 131 ATGTATTGGAAACTAGAGAGCTATTTTGTGTTTGGATTTTCTTCAAGTTCAAGGA 190  
 Db 1 MetTyrLeuGluThrArgAlaIlePheValPheTrpIlePheLeuGluValGlnGly 20  
 QY 191 ACCAAAGATATCTCCATACATATACCATTCGAACTAAAGACATAGACATCCCCCA 250  
 Db 21 ThrIleAspIleSerIleAsnIleTyrHisSerGluThrIleAspIleAspAsnProPro 40  
 QY 251 AGAATGAAACAACTGAAAGTACTGAAAAAATGTACAAAATCTCACTATGACGCAATA 310  
 Db 41 ArgAsnGluThrThrGluSerThrGluIleMetTyrIleSerMetThrMetArgIle 60  
 QY 311 TTTCGATTGGCAAGCATCGAACAAAGATCCGCATTTTCCCAAGGGGGTTAAAGTC 370  
 Db 61 PheAspLeuAlaIleYsHisArgThrIleArgSerAlaPhePheProThrGlyValIleVal 80  
 QY 371 TGTCCACAGGAATCCATGAAACAGATTTTACAGCTTTCAGCTTTCAGCTTATAGATTGAG 430  
 Db 81 CysProGlnGluSerMetIleGlnIleLeuAspSerLeuGlnAlaTyrTyrArgLeuArg 100  
 QY 431 GTGTCTCAGGAAGCAGTATGGGAAGCATATCGGATCTTCTGGATCCGATCCCTGACACA 490  
 Db 101 ValCysGlnGluAlaValTrpGluAlaTyrArgIlePheLeuAspArgIleProAspThr 120  
 QY 491 GGGGAATATCAGGACTGGGTGAGCATCTGCCAGCAGGAGACCTTCTGCTCTTTGACATT 550  
 Db 121 GlyGluTyrGlnAspTrpValSerIleCysGlnGlnGluThrPheCysLeuPheAspIle 140  
 QY 551 GCAAAAACTTCCAGCAATTTCCAGGAGCAGCTGGATCTTCTCAGCAGAGATAAAGAC 610  
 Db 141 GlyIleAsnPheSerAsnSerGlnGluHisLeuAspLeuLeuGlnArgIleIleGln 160  
 QY 611 AGAAGTTCCCTGCAGAGAAAGATGAATATCTGCAGAGAGACATTTGGAGAGCCTGGT 670  
 Db 161 ArgSerPheProAspArgIleAspGluIleSerAlaGluIleGluIleGluIleGluIle 180  
 QY 671 GAAACCATGCTCATTTTCAACAGCAGCATTCATTTCAAGAGCTTGGGCGATTTCTAAGAA 730  
 Db 181 GluThrIleValIleSer--Thr-----  
 QY 731 AACCTCAGAGAGCAAAATTCAGATCTTGCACAGCTCTCAGTTGGGCTTTCCCTCTCA 790  
 Db 188 -----AspValAlaAsnValSerLeuGlyProPheProLeuT 200  
 QY 791 CTCTCATGACACCTCTCTCAATGAATTCGATAATACACTCAACACACACAGATGC 850  
 Db 200 hrProAspThrLeuLeuAsnGluIleLeuAspAsnThrLeuAsnAspThrIleIleSerP 220  
 QY 851 CTAAACAGAGAGAAACAGATTCGCTGTGTGGAGGAGAGGCTGGAGTCAGGG 910  
 Db 220 roThrThrGluArgGluThrGluPheAlaValLeuGluGluGluArgValGluLeuSerV 240  
 QY 911 TCTCTCTGTAAACAGAGCTTCAAGGAGAGCTCGCTCAGCTCCAGCTCCGCTCCCATATTAC 970  
 Db 240 alSerLeuValAsnGlnIlePheIleAlaGluLeuAlaAspSerGlnSerProTyrTyrG 260  
 QY 971 AGAGCTAGAGAGAAAGTCCCACTTCAGATGCAAAAGATATTTAAGAAATTTCCAGGAT 1030  
 Db 260 InGluLeuAlaGlyLysSerGlnLeuGlnMetGlnIlePheIlePheIlePheIlePheIle 280  
 QY 1031 TCAAAAAATCCATGTGTTAGATTAGACCAAGAAAGAAAGAAAGATGGCTCAAGCTCCA 1090  
 Db 280 helYsIleHisValLeuGlyPheArgProIleIleGluIleAspGlySerSerSert 300  
 QY 1091 CAGAGATGCAACTTACGGCCATCTTTAAGAGACACAGTCGACAGACCAAGCCCTGCAA 1150  
 Db 300 hrGluMetGlnLeuThrAlaIlePheIleIlePheIleIleIleIleIleIleIleIleIle 320  
 QY 1151 GTGACCTCTGCTTTTGTATTCCCAAAATTTCAAGATGAGGAGTCTATCATGGAACCA 1210  
 Db 320 erAspLeuLeuSerPheAspSerAsnIleGluSerGluGluValTyrHisGlyThrM 340

QY 1211 TGGAGGAGCAAGCAACAGAAATCTATCTACAGCTACAGACTCAAAAGGCTGATCA 1270  
 Db etGluGluAspLysGlnProGluLileThyLeuThrAlaThrAspLeuLysArgLeuIleS 360  
 QY 1271 GCAAGCACTAGAGAGAGACAATCTTTGGATGTGGGACAAATTCAGTTCTACTGATGAAA 1330  
 Db erLysAlaLeuGluGluGlnSerLeuAspValGlyThyIleGlnPheThrAspGluI 380  
 QY 1331 TTGCTGGATCACTGCCAGCTTTGGTCTGACACCAATCAGAGCTGCCACATCTTTTG 1390  
 Db leaLagLysLeuProAlaPheGlyProAspThrGlnSerGluLeuProThrSerPheA 400  
 QY 1391 CTGTTTAACAGAGAGTCTACTTTGAGTCCAGAACTCTCTCTGTCGACCCACCTTG 1450  
 Db laValIleThrGluAspAlaThrLeuSerProGluLeuProValGluProGlnLeuG 420  
 QY 1451 AGACATGGAGGAGGAGGAGCATGTCTACCTGACACTCTTTGGTCTCCACTCTGCTATGG 1510  
 Db luThrValAspGlyAlaGluHisGlyLeuProAspThrSerTyrSerProAlaMetA 440  
 QY 1511 CCTTACCTCTCCTGTCAGAGCTCCACCTTTCTTTATGGCATCAAGCATCTCTCTGA 1570  
 Db laSerThrSerLeuSerGluAlaProProPhePheMetAlaSerSerIlePheSerLeuT 460  
 QY 1571 CTGATCAGGACACACAGATACAAATGGCCACTGACACAAATCTAGTACAGGGCTCA 1630  
 Db hrAspGlnGlyThrThrAspThrMetAlaThrAspGlnThrMetLeuValProGlyLeuT 480  
 QY 1631 CCATCCCAACAGTATTTCTGCAATCAGCCAACTGGCTCTGGGAATTCACATCCAC 1690  
 Db hrIleProThrSerAspTyrSerAlaIleSerGlnLeuAlaLeuGlyIleSerHisProp 500  
 QY 1691 CTGCATCTTCAGATCACAGCCGATCAAGTGCAGGTCGCGAGATATGTCACACCTAG 1750  
 Db toAlaSerSerAspAspSerArgSerSerAlaGlyGlyGluAspMetValArgHisLeuA 520  
 QY 1751 ATGAATGGATCTGTCTGCACTCTGCCCATCTGAGGTACAGAGCTCAGACCAATATG 1810  
 Db spGluMetAspLeuSerAspThrProAlaProSerGluValProGluLeuSerGluTyrV 540  
 QY 1811 TTTCTGCCCATCATTTCTGAGGATACCACTCTGTCTCAGCTTTACAGTATATCA 1870  
 Db alSerValProAspHisPheLeuGluAspThrThrProValSerAlaLeuGlnTyrIleT 560  
 QY 1871 CCATAGTTCTATGACCATTTGCCCAAGCGCGAGAGCTGTGCTGTCTTCTGCTCTGC 1930  
 Db hrThrSerSerMetThrIleAlaProLysGlyArgGluLeuValPhePheSerLeuA 580  
 QY 1931 GTGTTGCTAACATGCTCTCCAGCACTGTTCACAGAGCTCTCTGAGTACCGAG 1990  
 Db rgValAlaAsnMetAlaPheSerAsnAspLeuPheAsnLysSerSerLeuGluTyrArgA 600  
 QY 1991 CTCTGGAGCAACAATTCACAGCTGTGCTGCTCATATCTACGATCCATCTTACAGGAT 2050  
 Db laLeuGluGlnPheThrGlnLeuValProTyrLeuArgSerAsnLeuThrGlyP 620  
 QY 2051 TTAAAGCACTTGAATTAATTAATTCAGAAACGGAGTGTGATGTGATAGCAAAATCA 2110  
 Db helyGlnLeuGluIleLeuAsnPheArgAsnGlySerValIleValAsnSerLysMetL 640  
 QY 2111 AGTTTGTCAAGTCTGCGGTATACCTTACCAAGCTGTGACGGGGCTTGGAGGATT 2170  
 Db ysPheAlaLysSerValProTyrAsnLeuThrLysAlaValHisGlyValLeuGluAspP 660  
 QY 2171 TTGTTTCTGTGAGCCCAACCACTCCATCTGGAATAGACAGCTACTCTTCAACATTG 2230  
 Db heArgSerAlaAlaAlaGlnGlnLeuHisLeuGluIleAspSerTyrSerLeuAsnIleG 680  
 QY 2231 AACCACTGATTAAGCAGATCCCTGCAAGTTCTGCGCTGCGGGAATTTGCCAATGTG 2290  
 Db luProAlaAspGlnAlaAspProCysLysPheLeuAlaCysGlyGluPheAlaGlnCysV 700

QY 2291 TAAAGAAACGAGTCTGAGAACCGAGTGTCTGCTGCAACACAGGATATGACAGCCAGG 2350  
 Db alLysAsnGluArgThrGluGluAlaGluCysArgCysLysProGlyTyrAspSerGlnG 720  
 QY 2351 GAGCCCTGGACCGTCTGGAACCGAGCTCTGTGGCCCTGGCACAAGAGSNAATGCCAGTCC 2410  
 Db lySerLeuAspGlyLeuGluProGlyLeuCysGlyLeuAlaGlnArgAsnAlaArgSer 739  
 QY 2411 TCCAGGGAAGGAGCTCCATGTCAGGTTGCCAGATCACTCTGAAATCAAGCATACAAA 2470  
 Db SerArgGluArgGluLeuHisAla--ValProAspHisSerGluAsnGlnAlaTyrIlyst 759  
 QY 2471 CTAGTGTAAAGTTCCTCAAAATCAACAAATAACAGG 2509  
 Db hrSerValLysSerSerLysIleAsnLysIleThrArg 771

RESULT 5

ADA14867  
 ID ADA14867 standard; protein; 771 AA.  
 XX

AC ADA14867;  
 XX

DT 06-NOV-2003 (first entry)  
 XX

DE Human variant IPMC 150 isoform A.  
 XX

KW human; IPMC 150 isoform A; gene therapy;  
 KW interphotoreceptor matrix component; IPMC; ocular disorder;  
 KW macular degeneration; photoreceptor death; retinal detachment.  
 XX Homo sapiens.  
 OS

PN US2002160954-AI.  
 XX

PD 31-OCT-2002.  
 XX

PF 08-NOV-2001; 2001US-00007270.  
 XX

PR 29-OCT-1998; 98US-00183972.  
 XX

PR 29-OCT-1999; 99US-00430195.  
 XX

PA (TOWA ) UNIV IOWA RES FOUND.  
 XX

PI Hageman GS, Kuehn MH;  
 XX

DR WPI: 2003-238235/23.  
 XX

DR N-PSDB; ADA14866.  
 XX

PT New isolated or recombinant interphotoreceptor matrix component  
 PT polynucleotide and polypeptide, useful for diagnosing, preventing,  
 PT treating or prognosticating ocular disorders, e.g. macular degeneration  
 or retinal detachment.  
 PS Claim 8; Page 66-68; 76pp; English.  
 XX

CC The invention relates to an isolated or recombinant interphotoreceptor  
 CC matrix component (IPMC) polynucleotide. Also disclosed is a vector  
 CC comprising a promoter of an interphotoreceptor matrix component (IPMC)  
 CC gene operatively linked to the IPMC polynucleotide. The IPMC  
 CC polynucleotides, polypeptides and antibodies are useful for diagnosing,  
 CC preventing, treating or prognosticating ocular disorders, e.g. macular  
 CC degeneration, photoreceptor death or retinal detachment. They are also  
 CC useful for identifying a compound capable of modulating IPMC gene  
 CC expression in a cell. The present sequence represents the amino acid  
 CC sequence of the variant human interphotoreceptor matrix component, IPMC,  
 CC 150 isoform A.  
 XX

SQ Sequence 771 AA;

Alignment Scores:  
 Pred. No.: 0 Length: 771  
 Score: 3862.00 Matches: 769  
 Percent Similarity: 96.74% Conservative: 2

|    |      |  |      |
|----|------|--|------|
| QY | 1151 | GTGACCTCTCTGCTTTTGTGATTCACAAAAATTGAAAGTGGAGAACTCTATCATGGACCA   | 1210 |
| Db | 320  | erAspLeuLeuSerPheAspSerAsnLysAlleGleuSerGluValTyRHsglyThrM     | 340  |
| QY | 1211 | TGGAGGAGGACAGCACACCAAGAAATCTATCTCAGCTCAGACTCAGAAAGCTGATCA      | 1270 |
| Db | 340  | etGluGluAspLysGlnProGluIleTyLeuThrAlaThrAspLeuLysArgLeuIleS    | 360  |
| QY | 1271 | GCAAGCACTAGACGAGAAACAATCTTTGGATGTGGGACAAATTCAGTTCACGTATGAAA    | 1330 |
| Db | 360  | erLysAlaLeuGluGluGlnSerLeuAspValGlyThrIleGlnPheThrAspGluI      | 380  |
| QY | 1331 | TTGCTGGATCACTGCGAGCCTTTGGTCTCTGACACCCAAATCAGAGCTGCCACATCTTTTG  | 1390 |
| Db | 380  | leAlaGlySerLeuProAlaPheGlyProAspThrGlnSerGluLeuProThrSerPheA   | 400  |
| QY | 1391 | CTGTATTAAACAGAGATGCTACTTTGTAGTCCAGAACTTCCTCTGTTGAACCCAGCTTG    | 1450 |
| Db | 400  | laValIleThrGluAspAlaThrLeuSerProGluLeuProValGluProGlnLeuG      | 420  |
| QY | 1451 | AGACGTGACGGAGCAGACGATGCTACTACTGACACTTCTGTCTCCACCTCATATGG       | 1510 |
| Db | 420  | luThrValAspGlyAlaGluHisGlyLeuProAspThrSerTpSerProProAlaMetA    | 440  |
| QY | 1511 | CCCTTACCTCCCTCTCAGAGGCTCCACTTTCTTTATGGCATCAAGCATCTTCTCTCTGA    | 1570 |
| Db | 440  | laSerThrSerLeuSerGluAlaProPhePheMetAlaSerSerIlePheSerLeuT      | 460  |
| QY | 1571 | CTGATCAAGGACCAACAGATCAATGCCACTGACACAGAAATGCTAGTACACAGGCTCA     | 1630 |
| Db | 460  | hrAspGlnGlyThrThrAspThrMetAlaThrAspGlnThrMetLeuValProGlyLeuT   | 480  |
| QY | 1631 | CCATCCCAACCACTGATATTCTTGCATCAGCCAACTGGCTCTGGGAATTTCACTCCAC     | 1690 |
| Db | 480  | hrIleProThrSerAspTyrrSerAlaIleSerGlnLeuAlaLeuGlyIleSerHisProp  | 500  |
| QY | 1691 | CTGCATCTTCAGATGACAGCCGATCAGTGCAGGTGGCGAGATATGTCAGACACTAG       | 1750 |
| Db | 500  | roAlaSerSerAspAspSerArgSerSerAlaGlyGlyLeuAspMetValArgHisLeuA   | 520  |
| QY | 1751 | ATGAATATGATCTGTCTGACACTCTGCGCCATCTGAGTACACAGACTCAGCGATATG      | 1810 |
| Db | 520  | spGluMetAspLeuSerAspThrProAlaProSerGluValProGluLeuSerGluTyrv   | 540  |
| QY | 1811 | TTTCTGTGCCACATCATTTCTGGAGGATACCATCTCTGCTCTCAGCTTTACAGTATATCA   | 1870 |
| Db | 540  | aIleSerValProAspHisPheLeuGluAspThrThrProValSerAlaLeuGlnTyrlleT | 560  |
| QY | 1871 | CCATAGTTCTATACACATNTGCCCAAGGCGGAGCTGCTAGTGTCTTCAGTCTGC         | 1930 |
| Db | 560  | hrThrSerSerMetThrIleAlaProLysGlyArgGluLeuValPhePheSerLeuA      | 580  |
| QY | 1931 | GTGTTGCTTAAATGGCTCTTCCACGACCTGTTCAACAGAGCTCTCTGGAGTACCGAG      | 1990 |
| Db | 580  | rgValAlaAsnMetAlaPheSerAsnAspLeuPheAsnLysSerSerLeuGluTyrrArgA  | 600  |
| QY | 1991 | CTCTGGAGCAACAATTCACACAGCTGCTGGTCCATATCTACGATCCATCTTACAGAT      | 2050 |
| Db | 600  | laLeuGluGlnGlnPheThrGlnLeuLeuValProTyrrLeuArgSerAsnLeuThrGlyp  | 620  |
| QY | 2051 | TTAAGCAACTTGAATACTTAACCTTCAGAAACGGAGGTGATTTGTAATAGCAAAATGA     | 2110 |
| Db | 620  | heliysGlnLeuGluIleLeuAsnPheArgAsnGlySerValIleValAsnSerLysMetL  | 640  |
| QY | 2111 | AGTTTGTCTAGTCTGTGCGGTATTAACCTCACAAGGCTGTGCACGGGCTCTTGAGATT     | 2170 |
| Db | 640  | ysPheAlaLysSerValProTyrrAsnLeuThrLysAlaValHisGlyValLeuGluAspP  | 660  |
| QY | 2171 | TTCTGTTCTGTGCGGCCCAACAACCTCCATCTCGAAATAGACAGCTACTCTCTCAACATTG  | 2230 |
| Db | 660  | heArgSerAlaIleAlaGlnGlnLeuHisGluGluIleAspSerTyrrSerLeuAsnIleG  | 680  |

CC identifying a compound capable of modulating IPMC gene expression

XX

SQ Sequence 771 AA;

Alignment Scores:

|                        |         |               |     |
|------------------------|---------|---------------|-----|
| Pred. No.:             | 0       | Length:       | 771 |
| Score:                 | 3862.00 | Matches:      | 769 |
| Percent Similarity:    | 96.74%  | Conservative: | 2   |
| Best Local Similarity: | 96.49%  | Mismatches:   | 0   |
| Query Match:           | 65.74%  | Indels:       | 26  |
| DB:                    | 7       | Gaps:         | 1   |

US-10-007-270-1 (1-3330) x ABR42354 (1-771)

|    |     |   |      |
|----|-----|---|------|
| QY | 131 | ATGTATTTCGAACCTAGAGAGCTATTTTGTGTTTTTGGATTTTCTCCAGTTCAAGGA     | 190  |
| Db | 1   | MetTyLeuGluThrArgargalailePheValPheTrpIlePheLeuGlnValGlnGly   | 20   |
| QY | 191 | ACCAAGAATATCTCATTAACATATACCATTCGAACTTAAAGACATAGACAATCCCCCA    | 250  |
| Db | 21  | ThryasplleSerlelelanilelyrhiserGluThrLysAspIleAspAsnProPro    | 40   |
| QY | 251 | AGAAATGAACAACTGAAAGTACTGAAAATGTCAAAATGTCAACTATGAGACGAATA      | 310  |
| Db | 41  | ArgAsnGluThrThrGluSerThrGluLysMetTyTyLysMetSerThrMetArgIle    | 60   |
| QY | 311 | TTCGATTTCGCAAGCATCGAACAAAGATCCGATTTTCCCAACGGGGTTAAAGTC        | 370  |
| Db | 61  | PheAspLeuAlalyshIsargThrlyArgSerAlaPhePheProThrGlyLysVal      | 80   |
| QY | 371 | TGTCCACAGGAATCCATGAACACAGATTTTACAGCTTTCAGCTTATTATAGATTTCGA    | 430  |
| Db | 81  | CysProGlnGluSerMetLysGlnIleuAspSerLeuGlnAlaTyTyArgLysUrg      | 100  |
| QY | 431 | GTGTGTGAGGAAGCATGTGGGAAGCATATCGATCTTCTTCGATCGCATCCCTGACACA    | 490  |
| Db | 101 | ValCysGlnGluAlaValTrpGluAlaTyTyArgIlePheLeuAspArgIleProAspThr | 120  |
| QY | 491 | GGGGAATATCAGGACTGGGTGAGCATCTGCCAGCAGGAGACCTTCGGCTCTTTGACTT    | 550  |
| Db | 121 | GlyGluTyGlnAspTrpValSerIleCysGlnGlnGluThrPheCysLeuPheAspIle   | 140  |
| QY | 551 | GGAAAAACTTCAGCAATTTCCCGAGCAGCCTCGATCTTCTCCAGCAGAGATAAAACAG    | 610  |
| Db | 141 | GlyLysAsnPheSerAsnSerGlnGluHisLeuAspLeuLeuGlnIleArgIleLysGln  | 160  |
| QY | 611 | AGAGTTTCCCTGACAGAAAGATGAAATATCTGCAGAGAAGACATGGGAGAGCCTGGT     | 670  |
| Db | 161 | ArgSerPheProAspArgLysAspGluIleSerAlaGluLysThrLeuGluProGly     | 180  |
| QY | 671 | GAACCATTTGCATTTCAACAGCAATCTACATTTCAAGACTTGGGCAGTATCTTAGAA     | 730  |
| Db | 181 | GluThrIleValIleSer---Thr-----                                 | 187  |
| QY | 731 | AACCCCTCAGAGAGCAAAATTCAGATTTGCCAAGCTCTCACTTGGGCTTTCCCTCTCA    | 790  |
| Db | 188 | -----AspValAlaAsnValSerLeuGlyProPheProLeuTrp                  | 200  |
| QY | 791 | CTCCTGATGACACCTCTCAATGAATTTCTCGATAATATACATCTCAACAGACACCAAGATC | 850  |
| Db | 200 | hrProAspAspThrLeuLeuAsnGluIleLeuAspAsnThrLeuAsnAspThrLysMetP  | 220  |
| QY | 851 | CTACAAGAAAAGAGAAAAGAAATTCGTGTGTTGGAGGAGCAGAGGTGGAGCTCAGCG     | 910  |
| Db | 220 | roThrThrGluAspGluThrGluPheAlaValLeuGluGlnArgValGluLeuLeuSerV  | 240  |
| QY | 911 | TTCTCTGGTAAACACAGAAGTTCAAGCAGAGCTCGCTGACTCCAGTCCCCCATATTACC   | 970  |
| Db | 240 | alSerLeuValAsnGlnLysPheLysAlaGluLeuAlaAspSerGlnSerProTyTyG    | 260  |
| QY | 971 | AGGACTAGCAGGAAGTCCCACTTCAGATGCAAAAGATATTATTAAGAACTTCAGAGAT    | 1030 |
| Db | 260 | InGluLeuAlaGlyLysSerGlnGlnMetGlnLysIlePheLysLysLeuProGlyP     | 280  |

QY 1031 TCAGAAATCCATGCTGTAGCATTTAGACCAAGAAAGAAAGATGCTCAAGCTCCA 1090  
 Db |||||  
 QY 280 heLysylleHisValLeuGlyPheArgProLysLysGlyLysAspGlySerSerSert 300  
 Db |||||  
 QY 1091 CAGAGATGCAACTTACGGCCATCTTTAAGAGACACAGTGAGCAAGCAAAAGCCCTCAA 1150  
 Db |||||  
 QY 300 hrGluMetGlnLeuThrAlaIlePheLysArgHisSerAlaGluAlaLysSerProAlaS 320  
 QY 1151 GTGACCTTCCTGCTTTTGATTCACAAAATTGAAAGTGAGGAAGTCTATCATCGAACCA 1210  
 Db |||||  
 QY 320 erAspLeuLeuSerPheAspSerAsnLysIleGluSerGluValThrHisGlyThrM 340  
 QY 1211 TCGAGGAGGACAGCAACCAATCTATCTCAGCTACAGCTCAAAAGGCTGATCA 1270  
 Db |||||  
 QY 340 etGluGluAspLysGlnProGluIleTyrLeuThrAlaThrAspLeuLysArgLeuIleS 360  
 QY 1271 GCAAAGCACTAGAGGAGAACAAATCTTTGGATGTGGGACAAATTCAGTTCACTGATGAA 1330  
 Db |||||  
 QY 360 erLysAlaLeuGluGluGlnSerLeuAspValGlyThrIleGlnPheThrAspGluI 380  
 QY 1331 TTGCTGATCACTGCCAGCTTTGCTCTGACACCCCAATCAGAGCTGCCACATCTTTTG 1390  
 Db |||||  
 QY 380 leaLagLysLeuProAlaPheGlyProAspThrGlnSerGluLeuProThrSerPheA 400  
 QY 1391 CTGTTATACAGAGGATGCTACTTTGAGTCCAGAACTTCTCTCTGAGCCAGCTTG 1450  
 Db |||||  
 QY 400 laValIleThrGluAspAlaThrLeuSerProGluLeuProValGluProGlnLeuG 420  
 QY 1451 AGACAGTGGACGGAGCAGCATGCTTACCTGACACTTCTGCTCCACCTGCTATGG 1510  
 Db |||||  
 QY 420 luthrValAspGlyAlaGluHisGlyLeuProAspThrSerTrpSerProAlaMetA 440  
 QY 1511 CCTCTACCTCCCTGCTCAGAGCTCCACTTTCTTTATGCGATCAAGCATCTTCTCTGA 1570  
 Db |||||  
 QY 440 laSerThrSerLeuSerGluAlaProPhePheMetAlaSerSerIlePheSerLeuT 460  
 QY 1571 CTGATCAAGGACACACAGATACAAATGGCCACTGACAGCATGCTAGTACAGGGCTCA 1630  
 Db |||||  
 QY 460 hrAspGlnGlyThrThrAspThrMetAlaThrAspGlnThrMetLeuValProGlyLeuT 480  
 QY 1631 CCATCCCAACAGTGATTAATCTGCAATCAAGCAACTGCTCTGGGAATTCACATCCAC 1690  
 Db |||||  
 QY 480 hrIleProThrSerAspTyrSerAlaIleSerGlnLeuAlaGlyIleSerHisProp 500  
 QY 1691 CTGCATCTTCAGATGACAGCCGATCAAGCTGAGTGCGAAGATATGCTCAGACACCTAG 1750  
 Db |||||  
 QY 500 roAlaSerSerAspAspSerArgSerSerAlaGlyGlyGluAspMetValArgHisLeuA 520  
 QY 1751 ATGAAATGGATCTGCTGACACTCTGCCCATCTGAGTACACAGCTCAGCGAATAG 1810  
 Db |||||  
 QY 520 spGluMetAspLeuSerAspThrProAlaProSerGluValProGluLeuSerGluTyrV 540  
 QY 1811 TTTCTGTGCCAGATCATTTCTGGAGGATACACACTCTGCTCAGCTTTACAGTATATCA 1870  
 Db |||||  
 QY 540 aLerValProAspHisPheLeuGluAspThrThrProValSerAlaLeuGlnTyrIleT 560  
 QY 1871 CCATCTAGTCTTATGACCATGCCCCCAAGCGCGAGCTGCTAGTGTCTTCAGTCTGC 1930  
 Db |||||  
 QY 560 hrThrSerSerMetThrIleAlaProLysGlyArgGluLeuValPhePheSerLeuA 580  
 QY 1931 GTGTGTCTAACATGCGCTTCTCCAAAGCACTGTTCAACAGAGCTCTCTGGAGTACCGAG 1990  
 Db |||||  
 QY 580 rgValAlaAsnMetAlaPheSerAsnAspLeuPheAsnLysSerSerLeuGluTyrArgA 600  
 QY 1991 CTCTGGAGCAACATTCACAGCTGCTGTTTCCATATCTACGATCCATCTTACAGAT 2050  
 Db |||||  
 QY 600 laLeuGluGlnGlnPheThrGlnLeuLeuValProTyrLeuArgSerAsnLeuThrGlyP 620  
 QY 2051 TTAAGCAACTTGAATCTTAATCTCAGAAACGGAGTGTGATGTGAATAGCAAAATCA 2110  
 Db |||||  
 QY 620 heLysGlnLeuGluIleLeuAsnPheArgAsnGlySerValIleValAsnSerLysMetL 640

QY 2111 AGTTTGCTAAGTCTGTGCGGTATACCTCTCCAGAGCTGTGACGCGGTCTTCGAGGATT 2170  
 Db |||||  
 QY 640 yPheAlaLysSerValProTyrAsnLeuThrLysAlaValHisGlyValLeuGluAspP 660  
 QY 2171 TTCTGTTCTGTGCGAGCCCAACAACTCCATCTCTGGAATAGACAGCTACTCTCTCAATTTG 2230  
 Db |||||  
 QY 660 heArgSerAlaAlaAlaGlnGlnLeuHisLeuGluLysAspSerTyrSerLeuAsnIleG 680  
 QY 2231 ACCAGCTGTATCAAGCAGATCCCTGCAAGTTCCTGCGCTGCGCGCAATTTGCCCAATCTG 2290  
 Db |||||  
 QY 680 luProAlaAspGlnAlaAspProCysLysPheLeuAlaCysGlyGluPheAlaGlnCysV 700  
 QY 2291 TAAAGACCAACGACTGAGGAGCGGAGTCTGCTGCAAAACCAAGATATGACCCAGG 2350  
 Db |||||  
 QY 700 alLysAsnGlnArgThrGluGluAlaGluCysArgCysLysProGlyTyrAspSerGlnG 720  
 QY 2351 GGAGCTGCGAGCTCTGGAACACAGGCTCTGTGGCCCTGTCGACCAAGAGGAATCGGAGTCC 2410  
 Db |||||  
 QY 720 lySerLeuAspGlyLeuGluProGlyLeuCysGlyLeuAlaGlnArgAsnAlaArgSer 739  
 QY 2411 TCCAGGAAAGGAGCTCCATGCGAGTTCGCGAGTTCGCAAAATCAAGCATACAAAA 2470  
 Db |||||  
 QY 740 SerArgGluArgGluLeuHisAla--ValProAspHisSerGluAsnGlnAlaTyrLysT 759  
 QY 2471 CTAGTGTAAAAGTTTCCAAAATCAACAAATTAACAGG 2509  
 Db |||||  
 QY 759 hrSerVal-LysSerSerLysIleAsnLysIleThrArg 771  
 RESULT 7  
 ADA14843  
 ID ADA14843 standard; protein; 719 AA.  
 XX  
 AC ADA14843;  
 XX  
 DT 06-NOV-2003 (first entry)  
 XX  
 DE Human interphotoreceptor matrix component, IPMC, 150 isoform B.  
 XX  
 KW human; IPMC 150 isoform B; gene therapy;  
 KW interphotoreceptor matrix component; IPMC; ocular disorder;  
 KW macular degeneration; photoreceptor death; retinal detachment.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..20  
 FT Region /note= "Signal sequence"  
 FT Protein 1..7  
 FT /note= "Not encoded by {seqid:3}"  
 FT 21..719  
 FT /note= "Mature IPMC 150 isoform B"  
 XX  
 PN US2002160954-A1.  
 XX  
 PD 31-OCT-2002.  
 XX  
 PF 08-NOV-2001; 2001US-00007270.  
 XX  
 PR 29-OCT-1998; 98US-00183972.  
 PR 29-OCT-1999; 99US-00430195.  
 XX  
 PA (IOWA ) UNIV IOWA RES FOUND.  
 XX  
 PI Hageman GS, Kuehn MH;  
 XX  
 DR WPI; 2003-238235/23.  
 DR N-PSDB; ADA14842.  
 XX  
 PT New isolated or recombinant interphotoreceptor matrix component  
 PT polynucleotide and polypeptide, useful for diagnosing, preventing,  
 PT treating or prognosticating ocular disorders, e.g. macular degeneration  
 PT or retinal detachment.  
 XX



PS Claim 8; Page 34-36; 76pp; English.

XX The invention relates to an isolated or recombinant interphotoreceptor  
 CC matrix component (IPMC) polynucleotide. Also disclosed is a vector  
 CC comprising a promoter of an interphotoreceptor matrix component (IPMC)  
 CC gene operatively linked to the IPMC polynucleotide. The IPMC  
 CC polynucleotides, polypeptides and antibodies are useful for diagnosing,  
 CC preventing, treating or prognosticating ocular disorders, e.g. macular  
 CC degeneration, photoreceptor death or retinal detachment. They are also  
 CC useful for identifying a compound capable of modulating IPMC gene  
 CC expression in a cell. The present sequence represents the amino acid  
 CC sequence of human interphotoreceptor matrix component, IPMC, 150 isoform  
 XX B.

SQ Sequence 719 AA;

Alignment Scores:

Pred. No.: 0 Length: 719  
 Score: 3611.00 Matches: 717  
 Percent Similarity: 87.67% Conservative: 1  
 Best Local Similarity: 87.55% Mismatches: 1  
 Query Match: 61.46% Indels: 100  
 DB: 6 Gaps: 2

US-10-007-270-1 (1-3330) x ADA14843 (1-719)

QY 131 ATGATTTGGAACTAGAGAGCTATTTTGGTATTTTGGATTTTCTCCAGTTCAAGA 190  
 DB 1 MetTyrLeuGluThrArgAlaIlePheValPheTyrPheLeuGluValGlnGly 20  
 QY 191 ACCAAGATATCTCCATTAACATATACCATCTTGAACCTAAGACATAGACATCCCCA 250  
 DB 21 ThrLys----- 22  
 QY 251 AGAATGAACAACGTAAAGTACTGAAAAAATGTACAAAATGTCAACTATGAGCAATA 310  
 DB 22 ----- 22  
 QY 311 TTCGATTTGGCAAGCATCGAACAAAGATCCGATTTTCCACGGGGTTAAAGTC 370  
 DB 22 ----- 22  
 QY 371 TGTCCACAGGAATCCATGAACAGATTTTAGACAGTCTTCAAGCTTATTATAGATTGAGA 430  
 DB 22 ----- 22  
 QY 431 GTGTCTCAGGAGCAGTATGGAAGCATATCGGATCTTCTGGATCGCATCCCTGCACACA 490  
 DB 23 ValCysGlnGluAlaValTyrPheGluAlaTyrArgIlePheLeuAspArgIleProAspThr 42  
 QY 491 GGGGAATATCAGGACTGGCTCAGCATCTGCCAGCAGCAGTCTTCTCCAGCAGAGATAAAGCAG 610  
 DB 63 GlyLysAsnPheSerAsnSerGlnGluHisLeuAspLeuLeuGlnGlnArgIleLeuGln 82  
 QY 611 AGAAGTTTCCCTCAGCAAGAAAGATGAATATCTGCAGAGAGACATTTGGGAGAGCTGGT 670  
 DB 83 ArgSerPheProAspArgLysAspGluIleSerAlaGluLysThrLeuGluGluProGly 102  
 QY 671 GAAACCATTTGCTATTTCAACAGCAATCTACATTTCAAGACTTGGGAGTATCTTAAGAA 730  
 DB 103 GluThrIleValIleSer--Thr----- 109  
 QY 731 AACCTCAGAGAGCAAAATTCAGATTTGCCAAGCTCTCATTGGGCGCTTTCCCTCTCA 790  
 DB 110 -----AspValAlaAsnValSerLeuGlyProPheProLeuT 122  
 QY 791 CTCCTGATGACACCTCTCAATGAATTTCTCGAATAATCTCAACTCAACAGCACCAAGATGC 850  
 DB 122 hrProAspAspThrLeuLeuAsnGluIleLeuAspAsnThrLeuAsnAspThrLysMetP 142

QY 851 CTACACACAG 910  
 DB 142 rofhrThrGluArgGluThrGluPheAlaValLeuGluGluGlnArgValGluLeuSerV 162  
 QY 911 TCTCTCTGGTAAACACAGAGTTCAGGCGAGAGCTCGCTGACTCCAGTCCCATATTATACC 970  
 DB 162 alSerLeuValAsnGlnLysPheLeuAlaGluLeuAlaAspSerGlnSerProTyrG 182  
 QY 971 AGGAGCTACAGAGAGAGTCCCACTTCAGATTCGCAAGAGATATTAGAAATCTTCAGAT 1030  
 DB 182 InGluLeuAlaGlyLysSerGlnLeuGlnMetGlnLysIlePheLysLysLeuProGlyP 202  
 QY 1031 TCAAAAAAATCCATGTTAGAGATTTAGACCAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1090  
 DB 202 helYsLysIleHisValLeuGlyPheArgProLysLysGluLysAspGlySerSerSert 222  
 QY 1091 CAGAGATGCACTTACGGCCATCTTTAAGACACACAGTCGAGAGCAAAAGAGAGAGAGAG 1150  
 DB 222 hrGluMetGlnLeuThrAlaIlePheLysArgHisSerAlaGluAlaLysSerProAlas 242  
 QY 1151 GTGACCTCTGCTCTTTGATTCCACAAATTTGAAGTGAAGAGAGAGAGAGAGAGAGAGAG 1210  
 DB 242 erAspLeuLeuSerPheAspSerAsnLysIleGluSerGluGluValTyrHisGlyThrM 262  
 QY 1211 TGGAGGAGACAGCAACACAGAGATCTATCTCAGAGTACAGAGCTCAAAAGAGCTGATCA 1270  
 DB 262 etGluGluAspLysGlnProGluIleTyrLeuThrAlaThrAspLeuLysArgLeuLys 282  
 QY 1271 GCAAGACACTAG 1330  
 DB 282 erLysAlaLeuGluGluGlnSerLeuAspValGlyThrIleGlnPheThrAspGluI 302  
 QY 1331 TTCTCGATCACTCCAGAGCTTTGCTGTCGACACCAATCAGAGTCCACACATCTTTTG 1390  
 DB 302 leaLagLysLeuProAlaPheGlyProAspThrGlnSerGluLeuProThrSerPheA 322  
 QY 1391 CTGTTAATACAGAGAGAGTCTTCTGAGTCCAGAGTCTCTCTCTGTAACCCAGCTTG 1450  
 DB 322 laValIleThrGluAspAlaThrLeuSerProGluLeuProValGluProGlnLeuG 342  
 QY 1451 AGCAGTGGAGCGGAG 1510  
 DB 342 luthrValAspGlyAlaGluHisGlyLeuProAspThrSerTrpSerProAlaMetA 362  
 QY 1511 CCTCTACCTCTCTCAGAGAGCTCCACCTTTCTTTATGTCATCAGAGCATCTCTCTCTGA 1570  
 DB 362 laSerThrSerLeuSerGluAlaProProPhePheMetAlaSerSerIlePheSerLeuT 382  
 QY 1571 CTGATCAAGGACACACAGATACAATGGGCCACTGACACAGACAATCTAGTACACAGGCTCA 1630  
 DB 382 hrAspGlnGlyThrThrAspThrMetAlaThrAspGlnThrMetLeuValProGlyLeuT 402  
 QY 1631 CCATCCCAACAGTGAATTTCTGCAATCAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1690  
 DB 402 hrIleProThrSerAspPyrSerAlaIleSerGlnLeuAlaLeuGlyIleSerHisProp 422  
 QY 1691 CTGCACTCTCAGATGACAGCCCATCAAGTGGAGAGATATGGTCCAGACACCTAG 1750  
 DB 422 roAlaSerSerAspAspSerArgSerSerAlaGlyGlyGluAspMetValArgHisLeuA 442  
 QY 1751 ATGAAATGGATCTGCTGCACTCTCTGCCCCATCTGAGGTACAGAGCTCAGCAATATG 1810  
 DB 442 spGluMetAspLeuSerAspThrProAlaProSerGluValProGlyLeuSerGluTyrV 462  
 QY 1811 TTCTGTCCTCAGATCATTTCTTGGAGGATACCATCTCTCTCTCAGCTTACAGTATATCA 1870  
 DB 462 alSerValProAspPhePheLeuGluAspThrProValSerAlaLeuGlnIleT 482  
 QY 1871 CCATGATCTTATGACCATTTGCCCCCAAGGGCGAGAGCTGGTAGTCTTCTCAGTCTGC 1930  
 DB 482 hrThrSerSerMetThrIleAlaProLysGlyArgGluLeuValPhePheSerLeuA 502



|    |  |
|----|--|
| PF | 08-NOV-2002; 2002MO-US036090.  |
| XX |  |
| PR | 08-NOV-2001; 2001US-00077270.  |
| XX |  |
| PA | (IOWA ) UNIV IOWA RES FOUND.   |
| XX |  |
| PI | Hageman GS, Kuehn MH;  |
| XX |  |
| DR | MPI; 2003-441440/41.   |
| XX |  |
| PT | New interphotoreceptor matrix proteins and polynucleotides, useful for     |
| PT | treating or preventing photoreceptor death or retinal detachment, or for   |
| PT | treating ocular disorders.   |
| XX |  |
| FS | Claim 6; Page 79; 105pp; English.  |
| XX |  |
| CC | The present sequence is the protein sequence of isoform B of novel human   |
| CC | interphotoreceptor matrix 150 (IPM 150), a member of the newly identified  |
| CC | interphotoreceptor matrix component (IPMC) gene family. The IPM 150 gene   |
| CC | is located on chromosome q31-q35, a region that also contains loci for     |
| CC | progressive bilateral choriorretinal atrophy, autosomal dominant Stargard, |
| CC | -like macular dystrophy, North Carolina macular dystrophy and Salla        |
| CC | disease. Members of the IPMC gene family have been identified in humans,   |
| CC | monkey, cow, goat, rabbit, dog, cat, pig, mouse and rat IPM. 2             |
| CC | Subfamilies are designated IPM 150 (or IMPG1) and IPM 200 (or IMPG2). The  |
| CC | invention provides IPM 150 and IPM 200 polynucleotides and polypeptides,   |
| CC | antibodies that specifically bind the polypeptides, and vectors            |
| CC | comprising the polynucleotides. A claimed method of treating or            |
| CC | preventing photoreceptor death or retinal detachment involves              |
| CC | administering an IPMC polynucleotide, polypeptide or antibody. Also        |
| CC | claimed is a method for identifying a compound capable of modulating IPMC  |
| CC | gene expression  |
| XX |  |
| SQ | Sequence 719 AA;   |

Alignment Scores:

| Pred. No.:             | Score:  | Length:       | 719 |
|------------------------|---------|---------------|-----|
| Percent Similarity:    | 3611.00 | Matches:      | 717 |
| Best Local Similarity: | 87.67%  | Conservative: | 1   |
| Query Match:           | 61.46%  | Mismatches:   | 1   |
| DB:                    | 7       | Indels:       | 100 |
|                        |         | Gaps:         | 2   |

US-10-007-270-1 (1-3330) x ABR42343 (1-719)

|    |     |  |     |
|----|-----|--|-----|
| QY | 131 | ATGTAATTGGAAACTAGACAGCTAATTTTGTGTTTTTGGAATTTTCCTCAAGTTCACAGA   | 190 |
| Db | 1   | MetTyrLeuGluThrArgargalailePheValpheTrpillePheLeuGlnValGlnGly  | 20  |
| QY | 191 | ACCAAAGATACTCCTTAACATATACCATTCTGAAACTRAAGACATAGCAATCCCCCA      | 250 |
| Db | 21  | ThrLys-----  | 22  |
| QY | 251 | AGAAATGAACAACACTGAAGTAGTCTGAAAAAATGTCAAAAATGTCAACTATGACAGCAATA | 310 |
| Db | 22  | -----  | 22  |
| QY | 311 | TTTCGATTTGGCAAGCATCGAACAAAAAGATTCGCATTTTTTCCCACGGGGGTAAAGTC    | 370 |
| Db | 22  | -----  | 22  |
| QY | 371 | TGTCCACAGGAATCCATGAACAGATTTTAGACAGTCTTCAAGCTTATTATAGATTGAGA    | 430 |
| Db | 22  | -----  | 22  |
| QY | 431 | GTGTCTCAGGAACAGTATGGGAAGCATATCCGATCTTTCTGGATCGCATCCCTGACACA    | 490 |
| Db | 23  | ValCysGlnGluAlaValTrpGluAlaTyArgliePheLeuaspargileProaspThr    | 42  |
| QY | 491 | GGGGAATATCAGACTGGGTGAGCATCTGCCAGCAGGAGACCTTCTGCCTCTTTGACATT    | 550 |
| Db | 43  | GlyGluTyrlGlnAspTrpValserileCysGlnGlnGluThrPheCysLeuPheAspile  | 62  |



KW mouse; IPMC 150 isoform A; gene therapy;  
 KW interphotoreceptor matrix component; IPMC; ocular disorder;  
 KW macular degeneration; photoreceptor death; retinal detachment.  
 OS Mus sp.  
 XX US2002160954-A1.  
 XX 31-OCT-2002.  
 XX 08-NOV-2001; 2001US-00007270.  
 XX 29-OCT-1998; 98US-00183972.  
 XX 29-OCT-1999; 99US-00430195.  
 XX (IOWA) UNIV IOWA RES FOUND.  
 XX Hageman GS, Kuehn MH;  
 DR WPI; 2003-238235/23.  
 DR N-PSDB; ADA14847.  
 XX New isolated or recombinant interphotoreceptor matrix component  
 PT polynucleotide and polypeptide, useful for diagnosing, preventing,  
 PT treating or prognosticating ocular disorders, e.g. macular degeneration  
 PT or retinal detachment.  
 XX Claim 8; Page 41-43; 76pp; English.  
 XX The invention relates to an isolated or recombinant interphotoreceptor  
 CC matrix component (IPMC) polynucleotide. Also disclosed is a vector  
 CC comprising a promoter of an interphotoreceptor matrix component (IPMC)  
 CC gene operatively linked to the IPMC polynucleotide. The IPMC  
 CC polynucleotides, polypeptides and antibodies are useful for diagnosing,  
 CC preventing, treating or prognosticating ocular disorders, e.g. macular  
 CC degeneration, photoreceptor death or retinal detachment. They are also  
 CC useful for identifying a compound capable of modulating IPMC gene  
 CC expression in a cell. The present sequence represents the amino acid  
 CC sequence of mouse interphotoreceptor matrix component, IPMC, 150 isoform  
 CC A.  
 XX SQ Sequence 798 AA;  
 Alignment Scores:  
 Pred. No.: 1.38e-222 Length: 798  
 Score: 2504.50 Matches: 527  
 Percent Similarity: 73.13% Conservative: 80  
 Best Local Similarity: 63.49% Mismatches: 180  
 Query Match: 42.63% Indels: 43  
 DB: 6 Gaps: 10  
 US-10-007-270-1 (1-3330) x ADA14848 (1-798)  
 QY 131 ATGTTTGGAACTAGACAGCTATTTTGGTTTGGATTTTTCCTCAAGTTCAGGA 190  
 Db 1 MetAsnPheGlnIleLysHisAlaIlePheValPheGlyIlePheLeuGlnValGlnGly 20  
 QY 191 ACCAAGATATCTCCATTAACATATACCATCTGTAAGTAAGACATAGACATCCOCCA 250  
 Db 21 IleLysAspThrSerIleLysIlePheSerSerGluIleLysAsnIleAspLysThrPro 40  
 QY 251 AGAATGAACTACTGAAAGTACTGAAATAATGTAACAAATGTCAACTATGACGAGATA 310  
 Db 41 ArgIleGluThrIleLysSerThrValHisLysValSerThrMetLysArgIle 60  
 QY 311 TTCGATTGGCAAGCATCGAACAAAGATCGGATTTTCCCAACGGGGGTAAAGTC 370  
 Db 61 PheAspLeuProLysLeuArgThrLysArgSerAlaLeuPhePro---AlaAsnIle 79  
 QY 371 TGTCACAGGAATCCATGAAACAGATTTTAGACAGCTCTTCACTTATATAGATTGAGA 430  
 Db 80 CysProGlnGluSerLeuArgGlnIleLeuAlaSerLeuGlnIleLysIleTyrArgLeuArg 99

QY 431 GTGTGTGACAGACAGTATGGGAAGCATATCGATCTTTCTGATCGATCCCTGACACA 490  
 Db 100 ValCysGlnGluValValTrpGluAlaTrpArgIlePheLeuAspArgIleProAspThr 119  
 QY 491 GGGGAATATCAGGACTCGGTGACATCTGCCAGCAGGAGACCTTCTGGCTTTTGACATT 550  
 Db 120 GluGluTyrGlnAspTrpValSerLeuCysGlnLysGluThrPheCysLeuPheAspIle 139  
 QY 551 GGAATAAATCTCAGCAATCCAGGAGACCTGATCTTCTCCAGCAGAGAAATAACAG 610  
 Db 140 GlyLysAsnPheSerAsnSerGlnGluHisLeuAspLeuGlnArgIleLysGln 159  
 QY 611 AGAAGTTTCCCTGACAAAGATGAAATATCTGACAGAGACATTTGGAGAGCCTGGT 670  
 Db 160 ArgSerPheProGlyArgLysAspGluThrAlaSerMetGluThrLeuGluAlaProThr 179  
 QY 671 GAAACCATTTGTCATTTCACAGCATCTACATTTCAAGACTTGGGAGATTTCTTAAGA 730  
 Db 180 GluAlaProValPro---Thr----- 186  
 QY 731 AACCTCAGAGAGCAAAATTCAGATGTTCACACGTCTCCTTGGCCCTTTCCCTCA 790  
 Db 187 -----AspValSerArgMetSerLeuGlyProPheProLeuP 199  
 QY 791 CTCCTGACACCCCTCTCAATGAATTCGATAATACACCTCAACAGCAACCAAGATGC 850  
 Db 199 roSerAspAspThrAspLeuLysGluIleLeuSerValThrLeuLysAspIleGlnLysP 219  
 QY 851 CTACACAGAAAGAAACA-----GAATTCGTCTGTGTGGAGAGC 892  
 Db 219 roThrThrGluSerLysThrGluProIleHisValSerGluPheSer-----SerGluG 237  
 QY 893 AGAGGTGGAGCTAGCCTCTCTGTGTAACAGAGATTCACAGGAGAGCTCGCTGACT 952  
 Db 237 IuLysValGluPheSerIleSerLeuProAsnHisArgPheLysAlaGluLeuThrAsnS 257  
 QY 953 CCCAGTCCCCATATTACAGAGCTAGCAGAAAGTCCCACTTCAGATGCAAAAGATAT 1012  
 Db 257 erGlySerProLysTrpGlnGluValGlnSerGlnLeuGlnLysIleP 277  
 QY 1013 TTAAGAACTTCCAGATTCACAAATAATCCATGTGTGTAGGATTTAGACCAAGAGAGAA 1072  
 Db 277 heLysLysLeuProGlyPheGlyGluIleArgValLeuGlyPheArgProLysLysGluG 297  
 QY 1073 AAGATGCTTCAAGCTCCACAGAGATGCAATTCACGCCATCTTTAAGAGACAGATGTCAG 1132  
 Db 297 IuAspGlySerSerSerThrGluIleGlnLeuMetAlaIlePheLysArgAspHisAlaG 317  
 QY 1133 AAGCAAAAGCCCTGCAAGTACCTCTCTTTTGTATTCACAAATAATGAAAGTGAGG 1192  
 Db 317 IuAlaLysSerProAspSerHisLeuLeuSerLeuAspSerAsnLysIleGluSerGluA 337  
 QY 1193 AGCTTCATCATGGAACCATCGAGGAGCAAGCAACAGAAATCTATCTCACAGCTACAG 1252  
 Db 337 rgIleHisGlyValIle---GluAspLysGlnProGluThrTrpLeuThrAlaThrA 356  
 QY 1253 ACCTCAAAAGCTCATCAGCAAGACATAGAGGAAGAAATCTTTGGATGTGGGACAA 1312  
 Db 356 spuLysLysLeuIleGlnLeuLeuAspGlyAspLeuSerLeuValGluGlyLysI 376  
 QY 1313 TTCAGTTCACTGATGAATTTGCTGATCACTGCCAGCCTTTGGTCTCTGACACCAATCAG 1372  
 Db 376 leProPheGlyAspGluValThrGlyThrLeu-----PheArgProValThrGluProA 394  
 QY 1373 AGCTCCCAACATTTTGTGTTTATAACAGAGATGCTACTTTGAGTCCAGAACTTCTCTC 1432  
 Db 394 spuLeuProLysProLeuAlaAspValThrIuAspAlaThrLeuSerProGluLeuProp 414  
 QY 1433 CTGTTTGAACCCAGCTTGACAGCTGACGAGAGAGCATGCTTACCT----- 1482  
 Db 414 heValGluProArgGluAlaValAspArgGluGlySerGluLeuProGlyMetSerS 434  
 QY 1483 -----GACACTTCTGGTCTCCACCTGCTACTGCGCTCTACTCTCGTGTGAGAGCTCAC 1537

```

Db 434 eLyAspSerSerTpsProProValSerAlaSerIleSerArgSerGluAenLeuP 454
QY 1538 CTTCTTTATGGCATCAGCATCTCTCTGACTGATCAAGGCACACAGATACATGG 1597
Db 454 roSerPhe---ThrProSerIlePheSerLeuAspAlaGlnSerProProLeuMetT 473
QY 1598 CCACGTACACAGATAGTAGTACAGGGCTCACATCCACCCACAGTATTTCTGCAA 1657
Db 473 hrThrGlyProThrAlaLeuIleProLysProThrLeuProThrIleAspTyrSerThrI 493
QY 1658 TCAGCAACTGGCTCGGAATTCACATCCACTGCTCATCTTCAGATGACAGCCCATCAA 1717
Db 493 leArgGlnLeuProLeuGluSerSerHisTrpProAlaSerSerSerAspArgGluLeuI 513
QY 1718 GTGCAAGTGGCGAAGATATGTCACACCATAGATGAATGATCTCTGACACTCTCTG 1777
Db 513 leThrSerSerHisAspThrIleArgAspLeuAspGlyMetAspValSerAspThrProA 533
QY 1778 CCCATCTGAGGTACACAGCTCAGCAATATGTTCTGTCGCCAGATCATTTCTTGAGG 1837
Db 533 laLeuSerGluIleSerGluLeuSerGlyTyrAspSerAlaSerGlyGlnPheLeuGluM 553
QY 1838 ATACCACTCTGCTCTCAGCTTTACAGTATATACCACTAGTTCTATGACCATGCCCCA 1897
Db 553 etThrThrProIleProThrValArgPheIleThrSerSerGluThrIleAlaThrL 573
QY 1898 AGGCCGACAGCTGGTAGTGTCTTCAGTCTGCGTGTCTTAACATGGCTTCTCCACG 1957
Db 573 ysGlyGlnGluLeuValPhePheSerLeuArgValAlaAenMetProPheSerTyrA 593
QY 1958 ACTTGTTCAACAGAGCTCTCGAGTACCGAGCTCTGAGCAACATTCACAGCTGC 2017
Db 593 sPLeuPheAenLysSerSerLeuGluTyrGlnAlaLeuGluGlnArgPheThrAspLeuL 613
QY 2018 TGGTTCCATATCTACATCCCAATCTTACAGGATTTAAGCACTTGAATTAATTAAC 2077
Db 613 euValProTyrLeuArgSerAenLeuThrGlyPheLysGlnLeuGluIleLeuSerPheA 633
QY 2078 GAACGGAGTGTGATTGTGAATAGCAAAATGAAGTTGCTTAAGTCTGTCGCGTATAAC 2137
Db 633 rGAsnGlySerValIleValAAsnSerLysValArgPheAlaLysAlaValProTyrAAsnL 653
QY 2138 TCACAGAGGTGTCACGCGGTCTGAGGATTTCTGCTGTCGAGCCACCACTCC 2197
Db 653 euThrGlnAlaValArgGlyValLeuGluAspLeuArgSerThrAlaAlaGlnGlyLeuA 673
QY 2198 ATCTGGAATPAGACAGTACTCTCTCAACATTGAACCAAGCTGATCAAGCAGATCCCTGCA 2257
Db 673 snLeuGluIleGluSerTyrSerLeuAspIleGluProAlaAspGlnAlaAspProCysL 693
QY 2258 AGTTCCTGGCTGCGCGAATTTGCCATGTGTAAAGAACGACGAGTCTGAGAGCGG 2317
Db 693 ysLeuLeuAspCysGlyLysPheAlaGlnCysValLysAsnGluTrpThrGluGluAlaG 713
QY 2318 AGTGTGCTGCAAAACAGGATATGACAGCGGAGCGCTGAGCGTCTGGAACAGGCC 2377
Db 713 luCysArgCysArgGlnGlyHisGluSerHisGlyThrLeuLeuAspTyrGlnThrLeuAAsnL 733
QY 2378 TCTGTGCTGCTGCAAAAGGAATGCGAGGTCTCCAGGAAGAGGAGCTCCATGAGGT 2437
Db 733 euCysProGly---LysThrCysValAlaGlyArgGluGlnAlaThrProCysArgP 752
QY 2438 TGCAGATCACTCTGAATCAAGCATACAAATAGTGTAAAGTTCGAAATCAAC 2497
Db 752 roProAspHisSerThrAAsnGlnAlaGlnGluProGlyValLysLysLeu-----ArgG 770
QY 2498 AAAATACAGGTAATCAGTAAAGAAATTTCTGAATTACTGACCGTACGATATCAAGAT 2557
Db 770 lngLnaenLysValValLysLysArgAAsnSerLysLeuSerAlaIleGlyPheGluGluP 790
QY 2558 TTAACCATCAAGATTGGAAGAAAT 2583

```

Db 790 heGluAspGlnAspTrpGluGlyAsn 798

# RESULT 10

ABR42345

ID ABR42345 standard; protein; 798 AA.

XX ABR42345;

XX 11-AUG-2003 (first entry)

DE Mouse interphotoreceptor matrix IPM 150, isoform A.

KW Mouse; interphotoreceptor matrix; IPM 150; IPMC; receptor; ophthalmological; gene therapy.

XX Mus sp.

Key Location/Qualifiers

Domain 71..80

/note= "N-terminal domain"

Domain 95..125

/note= "conserved domain 1"

Domain 405..577

/note= "mucin-like domain"

Domain 576..657

/note= "conserved domain 2"

Domain 697..740

/note= "EGF-like domain"

XX WO2003039346-A2.

XX 15-MAY-2003.

XX 08-NOV-2002; 2002WO-US036090.

XX 08-NOV-2001; 2001US-00077270.

XX (IOWA ) UNIV IOWA RES FOUND.

XX Hageman GS, Kuehn MH;

XX WPT; 2003-441440/41.

XX DR N-PSDB; ACC57950.

XX New interphotoreceptor matrix proteins and polynucleotides, useful for treating or preventing photoreceptor death or retinal detachment, or for treating ocular disorders.

XX Claim 6; Page 82; 105pp; English.

The present sequence is the protein sequence of isoform A of novel mouse interphotoreceptor matrix 150 (IPM 150), a member of the newly identified interphotoreceptor matrix component (IPMC) gene family. IPMC gene family members have been identified in humans, monkey, cow, goat, rabbit, dog, cat, pig, mouse and rat IPM. 2 Subfamilies are designated IPM 150 (or IPM1) and IPM 200 (or IPM2). The invention provides IPM 150 and IPM 200 polynucleotides and polypeptides, antibodies that specifically bind the polynucleotides, and vectors comprising the polynucleotides. A claimed method of treating or preventing photoreceptor death or retinal detachment involves administering an IPMC polynucleotide, polypeptide or antibody. Also claimed is a method for identifying a compound capable of modulating IPMC gene expression

SQ Sequence 798 AA;

Alignment Scores:

Pred. No.: 1.38e-222 Length: 798

Score: 2504.50 Matches: 527

Percent Similarity: 73.13% Conservative: 80

Best Local Similarity: 63.49% Mismatches: 180

Query Match: 42.63% Indels: 43

DB: 7 Gaps: 10



[illegible]

|    |      |  |      |
|----|------|--|------|
| Db | 146  | erGluGluLysValGluPheSerLysSerLeuProAsnHisArgPheLysAlaGluLeu  | 166  |
| Qy | 947  | CTGACTCCCCAGTCCCATATTAACAGAGAGTACAGAGAAAGTCCCAACTTCAGATCGAA  | 1006 |
| Db | 166  | hrAsnSerGlySerProTyrTyrGlnGluLeuValGlyGlnSerGlnLeuGlnLeuGln  | 186  |
| Qy | 1007 | AGATATTTAAGAACTTCAGAGATTCAAAAATCCATGTGTGTAGGATTTAGACCAAGA    | 1066 |
| Db | 186  | ybtlePheLysLeuProGlyPheGlyGluIleArgValLeuGlyPheArgProLysL    | 206  |
| Qy | 1067 | AAGAAAAGATGGCTCAAGCTCCACAGAGATCCACTTACGGCCATCTTTAAGAGACAC    | 1126 |
| Db | 206  | ysGluLysPglySerSerSerThrGluIleGlnMetAlaIlePheLysArgAspH      | 226  |
| Qy | 1127 | GTCCGAAGCAAAAAGCCCTCAAGTCACTCCTGTCTTTTGATTCCAACAAAATTGAA     | 1186 |
| Db | 226  | isAlaGluAlaLysSerProAspSerHisLeuLeuSerLeuAspSerAsnLysIleGluS | 246  |
| Qy | 1187 | GTGAGAGAGTCTATCATGGAACCATAGAGAGGAGCAAGCAACCAAGAAATCTATCTCA   | 1246 |
| Db | 246  | erGluArgIleHisGlyValIle---GluaspLysGlnProGluThrTyrLeuThrA    | 265  |
| Qy | 1247 | CTACAGACTCAAAAGCGTATCAGCAAGCACTAGAGAGAGAAACAATCTTGGATGTGG    | 1306 |
| Db | 265  | laThrAspLeuLysLysLeuIleGlnLeuLeuAspGlyAspLeuSerLeuValGluG    | 285  |
| Qy | 1307 | GGACAATTCACTCACTGATGAAATGCTGGATCACTGCGAGCTTTGGTCTCGACACC     | 1366 |
| Db | 285  | LysLysIleProPheGlyAspGluValThrGlyThrLeu-----PheArgProValThrG | 303  |
| Qy | 1367 | AATCAGACTGCCCAATCTTTCTGTTTATAACAGAGAGATCTACTTTGAGTCAGAAC     | 1426 |
| Db | 303  | luProAspLeuProLysProLeuAlaAspValThrGluaspAlaThrLeuSerProGluL | 323  |
| Qy | 1427 | TTCTCTCTGTGTAACCCAGCTTGACAGAGTGGAGGAGGAGGAGCGTGTCTACCT---    | 1482 |
| Db | 323  | euProPheValGluProArgLeuGluAlaValAspArgGluGlySerGluLeuProGlyM | 343  |
| Qy | 1483 | -----GACATCTTTGGTCTCCACTGCTATGGCTCTACTCTCTCTCTCTCTCTCTCT     | 1531 |
| Db | 343  | etSerSerLysAspSerSerTyrSerProProValSerAlaSerIleSerArgSerGluA | 363  |
| Qy | 1532 | CTCCACTTTCTTTATGGCAFCAGCAATCTCTCTGTGATCATCAGGACCAACAGATA     | 1591 |
| Db | 363  | snLeuProSerPhe--ThrProSerIlePheSerLeuAspAlaGlnSerProProL     | 382  |
| Qy | 1592 | CAATGGCCACTGACCAACAATGCTAGTACAGGGCTCCACATCCCAACCAAGTATTATT   | 1651 |
| Db | 382  | euMetThrThrGlyProThrAlaLeuIleProLysProThrLeuProThrIleAspTyrS | 402  |
| Qy | 1652 | CTGCAAATCAGCCAACTGGCTCTGGGAATTCACATCCACTGTCATCTTCAGATGACGCC  | 1711 |
| Db | 402  | erThrIleArgGlnLeuProLeuGluSerSerHisTrpProAlaSerSerSerAspArgG | 422  |
| Qy | 1712 | GATCAAGTCAGCTGGCAGAGATAGTCAGACACTAGTAAATGGATCTGTGTGACA       | 1771 |
| Db | 422  | luLeuIleThrSerSerHisAspThrIleArgAspLeuAspGlyMetAspValSerAspT | 442  |
| Qy | 1772 | CTCTCTGCCCATCTCAGGTACAGAGCTCAGCGAATATCTTCTGTCGCCAGATCATTTCT  | 1831 |
| Db | 442  | hrProAlaLeuSerGluIleSerGluLeuSerGlyTyrAspSerAlaSerGlyGlnPheL | 462  |
| Qy | 1832 | TGGAGGATCACTCTCTCTACGCTTTTACAGTATATCACCATTAGTTCTATGACCAATTG  | 1891 |
| Db | 462  | euGluMetThrThrProIleProThrValArgPheIleThrThrSerSerGluThrIleA | 482  |
| Qy | 1892 | CCCCAAGGCCGAGAGCTGTGTAGTGTCTTCTAGTCTGGCTGTGTCTTAACATGGCCCTTC | 1951 |
| Db | 482  | laThrLysGlyGlnGluLeuValValPhePheSerLeuArgValAlaAsnMetProPheS | 502  |
| Qy | 1952 | CCACAGCACTGTTTCAACAGAGCTCTCTGGAGTACCGAGCTCTGGAGCAACAATTTCAC  | 2011 |
| Db | 502  | erTyrAspLeuPheAsnLysSerLeuGluTyrGlnAlaLeuGluGlnArgPheThrA    | 522  |

|           |  |   |      |
|-----------|--|---|------|
| Qy        | 2012   | AGCTGCTGGTTCATATCTACGATCCAAATCTTACAGGATTTAAAGCAACTTGAATACTCTA | 2071 |
| Db        | 522  | spLeuLeuValProTyrLeuArgSerAsnLeuThrGlyPheLysGlnLeuGluLeuLeuS  | 542  |
| Qy        | 2072   | ACTTCAGAACCGGAGTCTGATTTCTCAATAGCAAAATGAAGTTTGTCTAAGTCTCTGCCGT | 2131 |
| Db        | 542  | erPheArgAsnGlySerValIleValAsnSerLysValArgPheAlaLysAlaValProT  | 562  |
| Qy        | 2132   | ATAACCTCAACAGGCTGTGCACGGGCTCTGGAGGATTTTCGTTCTGTCGACGCCAAC     | 2191 |
| Db        | 562  | yrAsnLeuThrGlnAlaValArgGlyValLeuGluAspLeuArgSerThrAlaGlnG     | 582  |
| Qy        | 2192   | AACTCCATCTGGAATAGACAGCTACTCTCTCAACATTTGAACACAGCTGATCAAGCAGATC | 2251 |
| Db        | 582  | lyLeuAsnLeuGluIleGluSerTyrSerLeuAspIleGluProAlaAspGlnAlaAsp   | 602  |
| Qy        | 2252   | CCTGCAAGTTCTGSCCTCGCGGCGAATTTGCCCAATGCTGAAGAACGACGGAAGTGAAG   | 2311 |
| Db        | 602  | roCysLysLeuLeuAspCysGlyLysPheAlaGlnCysValLysAsnGluTrpThrGluG  | 622  |
| Qy        | 2312   | AAGCGGAGTGTCTGCAAAAACAGGATATGACGCCAGGAGGAGCTCGACGCTCTGGAAC    | 2371 |
| Db        | 622  | IuAlaGluCysArgGlnGlyHisGluSerHisGlyThrLeuAspTyrGlnThrL        | 642  |
| Qy        | 2372   | CAGGCTCTGTGGCCCTGTGGCACAAAGAAATCGAGGTCTCTCCAGGAAAGGAGCTCCAT   | 2431 |
| Db        | 642  | euAsnLeuCysProGlyLysThrCysValAlaGlyArgGluGlnAlaThrProC        | 661  |
| Qy        | 2432   | GCAGGTTGCCAGATCACTCTGAAATCAAGCATACAAACTAGTCTTAAAGTTCCAAA      | 2491 |
| Db        | 661  | ysArgProAspHisSerThrAsnGlnAlaGlnGluProGlyValLysLysLeu---      | 679  |
| Qy        | 2492   | ATCAACAAATAACAGGTAATCAGTAAAGAAATTTCTGAATTAATCTGACCGCTAGAATATG | 2551 |
| Db        | 680  | --ArgGlnGlnAsnLysValValLysLysArgAsnSerLysLysSerAlaIleGlyPheG  | 699  |
| Qy        | 2552   | AAGAAATTTAACCATCAAGATGGGAAGAAAT                               | 2583 |
| Db        | 699  | IuGlnPheGlu***GlnAspTrpGluGlyAsn                              | 709  |
| RESULT 12 |  |   |      |
| AD        | ADAL4850   |   |      |
| ID        | ADAL4850   | standard; protein; 466 AA.                                    |      |
| XX        | ADAL4850;  |   |      |
| AC        | AC   |   |      |
| XX        | XX   |   |      |
| DT        | 06-NOV-2003  | (first entry)   |      |
| DE        | Mouse  | interphotoreceptor matrix component, IPMC, 150 isoform D.     |      |
| DE        | mouse; IPMC 150 isoform D; gene therapy;                       |   |      |
| KW        | interphotoreceptor matrix component; IPMC; ocular disorder;    |   |      |
| KW        | macular degeneration; photoreceptor death; retinal detachment. |   |      |
| OS        | Mus sp.  |   |      |
| XX        | XX   |   |      |
| Key       | Location/Qualifiers  |   |      |
| FT        | Peptide  | 1..20   |      |
| FT        | FT   | /note= "Signal sequence"                                      |      |
| FT        | Protein  | 21..466   |      |
| FT        | FT   | /note= "Mature IPMC 150 isoform D"                            |      |
| FT        | FT   |   |      |
| PN        | US2002160954-A1.   |   |      |
| PD        | 31-OCT-2002.   |   |      |
| PF        | 08-NOV-2001; 2001US-00007270.                                  |   |      |
| PR        | 29-OCT-1998; 98US-00183972.                                    |   |      |
| PR        | 29-OCT-1999; 99US-00430195.                                    |   |      |
| XX        | (IOWA ) UNIV IOWA RES FOUND.                                   |   |      |

XX Hageman GS, Kuehn MH;  
 XX WPI; 2003-238235/23.  
 DR N-PSDB; ADAL4849.  
 XX  
 PT New isolated or recombinant interphotoreceptor matrix component  
 PT polynucleotide and polypeptide, useful for diagnosing, preventing,  
 PT treating or prognosticating ocular disorders, e.g. macular degeneration  
 PT or retinal detachment.  
 XX  
 XX Claim 8; Page 44-45; 76pp; English.  
 XX  
 CC The invention relates to an isolated or recombinant interphotoreceptor  
 CC matrix component (IPMC) polynucleotide. Also disclosed is a vector  
 CC comprising a promoter of an interphotoreceptor matrix component (IPMC)  
 CC gene operatively linked to the IPMC polynucleotide. The IPMC  
 CC polynucleotides, polypeptides and antibodies are useful for diagnosing,  
 CC preventing, treating or prognosticating ocular disorders, e.g. macular  
 CC degeneration, photoreceptor death or retinal detachment. They are also  
 CC useful for identifying a compound capable of modulating IPMC gene  
 CC expression in a cell. The present sequence represents the amino acid  
 CC sequence of mouse interphotoreceptor matrix component, IPMC, 150 isoform  
 CC D.  
 XX  
 SQ Sequence 466 AA;  
 Alignment Scores:  
 Pred. No.: 7,73e-111 Length: 466  
 Score: 1298.50 Matches: 309  
 Percent Similarity: 43.76% Conservative: 52  
 Best Local Similarity: 37.45% Mismatches: 99  
 Query Match: 22.10% Indels: 365  
 DB: 6 Gaps: 10  
 US-10-007-270-1 (1-3330) x ADAL4850 (1-466)  
 QY 131 ATGTATTTGGAACATGAGAGCTATTTTGGATTTTCCAGTTCAGGA 190  
 DB 1 MetAsnPheGlnIleLysHisAlaIlePheValPheGlyIlePheLeuGlnValGlnGly 20  
 QY 191 ACCAAGATATCTCCATTAAACATATACCATCTCTGAAACCTAAAGACATAGACATCCCCA 250  
 DB 20 ----- 20  
 QY 251 AGAATGAACAACTGAAAGTACTGAAAAATGTCAAAATGTCAACTATGAGACGAATA 310  
 DB 20 ----- 20  
 QY 311 TTCGATTTGGCAAGCATCGAACAAAAGATCGCATTTTCCCAACGGGGTTAAAGTC 370  
 DB 20 ----- 20  
 QY 371 TGTCCACAGGAATCCATGAACACAGATTTTACAGAGTCTTCAAGCTTATTATAGATTGAGA 430  
 DB 21 -----11leIys 22  
 QY 431 GTGTCTCAGGAGCAGTATGGAAGCATATCGCATTTCTGGATCCATCCCTGACACA 490  
 DB 23 ValCysGlnGluValTrpGluAlaTrpArgIlePheLeuAspArgIleProAspThr 42  
 QY 491 GGGGAATATCAGAGCTGGGTGAGCATCTGCAGCAGGAGACCTTCTGCTCTTTTCACATT 550  
 DB 43 GluGluTyrglnAspTrpValserLeuCysGlnLysGluThrPheCysLeuPheAspIle 62  
 QY 551 GGAATAATCTCAGCAATTCCTCAGGAGCAGCTGATTTCTCCAGCAGAGAAATAAAGCAG 610  
 DB 63 GlyLysAsnPheSerGlnGlnHisLeuAspLeuGlnGlnArgIleLysGln 82  
 QY 611 AGAGTTTCCCTGACAGAAAAGATGAATATCTGCAGAGAGACATTTGGGAGGCTGGT 670  
 DB 83 ArgSerPheProGlyArgLysAspGluThrAlaSerMetGluThrLeuGluAlaProThr 102

QY 671 GAAACCAATGTCAATTTCAACAGCAATCTACATTTTCAAGACTTGGGCAGTATTCTAAGAA 730  
 DB 103 GluAlaProValPro--Thr----- 109  
 QY 731 AACCTCAGAGAGCAATTTCAAGATGTTGCCAAGCTCTCACTTGGGCTTCCCTCTCA 790  
 DB 110 -----AspValSerArgMetSerLeuGlyProLeuP 122  
 QY 791 CTCTGATCACACCTCTCTCAATGAAATTTCCGATATACACTCAACCCAGCAAGATCC 850  
 DB 122 roSerAspAspThrAspLeuLysGluIleLeuSerValThrLeuLysAspIleGlnLysP 142  
 QY 851 CTACACAGAGAGAAACA-----GAATTCCTGTGTGGAGAGC 892  
 DB 142 roThrThrGluSerIleThrGluProIleHisValSerGluPheSer-----SerGluG 160  
 QY 893 AGAGGTGTCAGCTCAGGCTCTCTGTGTAACACAGAGTTCAGGCAGAGCTCGCTGACT 952  
 DB 160 luySValGluPheSerIleSerLeuProAsnHisArgPheLysAlaGluLeuThrAsnS 180  
 QY 953 CCAGTCCCATATTTACCAGAGCTAGCAGGAAAGTCCCAACTTCAGATGCAGAAAGATAT 1012  
 DB 180 ergLysSerProTyrglnGluLeuValGlyGlnSerGlnLeuGlnLeuGlnLysIleP 200  
 QY 1013 TTAAGAACTTCCAGGATTCAAAAATCCATGTGTGTAGATTAGACCAAGAAAGAAA 1072  
 DB 200 helYsLeuProGlyPheGlyIleArgValLeuGlyPheArgProLysLysGluG 220  
 QY 1073 AAGATGGCTCAAGCTCCACAGAGATGCAACTTACGGCATCTTTAAGACACACAGTGCAG 1132  
 DB 220 luyAspGlySerSerSerThrGluIleGlnLeuMetAlaIlePheLysArgAspHisAlaG 240  
 QY 1133 AGCAAAAAGCCCTGCAAGTGAACCTCTGTCTTTGATTCCAAACAAATTTGAAGTGAGG 1192  
 DB 240 luAlaLysSerProAspSerHisLeuLeuSerLeuAspSerAsnLysIleLysSerGluA 260  
 QY 1193 AAGCTTATCATGGAACCATGAGGAGGAGACAGCAACACAAATCTATCTCACAGCTACAG 1252  
 DB 260 rgIleHisGlyValIle---GluAspLysGlnProGluThrTyrglnThrAlaThrA 279  
 QY 1253 ACCTCAAAAGGCTCATCAACCAAGCATAGAGGAAGCAACATCTTTGGATGTGGGACAA 1312  
 DB 279 spleuLysLeuLleGlnLeuLeuAspGlyAspLeuSerLeuValGluGlyLysI 299  
 QY 1313 TCCAGTTCACCTGATGAAATTCGTGATCCTGCGAGCTTGGTCTCGACACCAATCAG 1372  
 DB 299 leProPheGlyAspGluValThrGlyThrLeu-----PheArgProValThrGluProA 317  
 QY 1373 AGCTGCCACATCTTTTGTGTTTAAACAGAGGATGCTACTTTGAGTCCAGAACTTCTC 1432  
 DB 317 spleuProLysProLeuAlaAspValThrGluAspAlaThrLeuSerProGluLeuProP 337  
 QY 1433 CTGTTGAACCCAGCTTGAGACAGTGGACGAGGAGCAGAGCATGGTCTACCTGACACTTCTT 1492  
 DB 337 heValGluProArgLeuGluAlaValAsp----- 346  
 QY 1493 GGTCTCCACCTGTATGGGCTCTACCTCCCTGTGACAGACTCCACCTTTCTTTATGGCAT 1552  
 DB 346 ----- 346  
 QY 1553 CAAGCATCTTCTCTGACTGATCAAGGCACACACAGATACATAGCCACCTGACACACAA 1612  
 DB 346 ----- 346  
 QY 1613 TGCTAGTACAGGCTCACCATCCACCAGTGTATTTCTGCAATCAGCAACTGGCTC 1672  
 DB 346 ----- 346  
 QY 1673 TGGGAATTCACATCCACCTGCATCTTCAGATGACAGCCGATCAAGTGCAGGTGGCGAAG 1732  
 DB 346 ----- 346  
 QY 1733 ATATGTCAGACACCTAGATGAATGGATCTGTCTGACACTCTCTGCCCCATCTGAGGTAC 1792



Db 346 ----- 346  
 QY 1793 CAGAGCTCAGGAATATGTTCTGTCCAGATGATTTCTTGGAGGTACCACTCTCTGTCT 1852  
 Db 346 ----- 346  
 QY 1853 CAGCTTTACAGTATATCACCAGTATGTTCTATGACCATGTCCTCCCAAGGCGGAGAGCTGG 1912  
 Db 346 ----- 346  
 QY 1913 TAGTGTCTTTCAGTCTGCGTGTCTAAGTCTGCTTCTCAACGACCTGTTCAACAAGA 1972  
 Db 346 ----- 346  
 QY 1973 GCTCTCTGAGTACGAGCTCTGGAGCAACAATTCACAGCTGCTGTTCCATATCTAC 2032  
 Db 346 ----- 346  
 QY 2033 GATCCAACTTACAGGATTAAGCAACTTGAATACCTTAACTTCAGAAACGGAGTGTGA 2092  
 Db 347 ----- 352  
 QY 2093 TTGTGAATAGCAAAATGAAGTTTCTAAGTCTGTGCGGTATAACCTTCAACAGGCTGTGC 2152  
 Db 352 eu----- 352  
 QY 2153 ACGGGTCTTGGAGGATTTTCGTTCTGTCGAGCCCAACAACCTCCATCTCGAATAGACA 2212  
 Db 352 ----- 352  
 QY 2213 GCTACTCTCAACATTGAACAGCTGATCAAGCAGATCCCTGCAAGTCTCTGGCTGG 2272  
 Db 353 ----- 366  
 QY 2273 GCGAATTTCCCAATGTAAGAACCAACGAGTCTGAGAGCGAGTCTGCTGCAAC 2332  
 Db 366 llyyepheAlaGlnCysVallysaenGluTrpThrGluGluAlaGluCysArgCysArgG 386  
 QY 2333 CAGGATATGACAGCGAGGAGCTGAGCGTCTGGAACAGGCTCTGTGGCCCTGGCA 2392  
 Db 386 lnglyHisGluSerHisGlyThrLeuAspTyGlnThrLeuAsnLeuCysProProGly- 405  
 QY 2393 CAAGGAATGCGAGTCTCCAGGAAAGGAGTCTCCATGAGTTCAGATTCATCAAGATT 2452  
 Db 406 --LysThrCysValAlaGlyArgGluGlnAlaThrProCysArgProThrAspHisSerT 425  
 QY 2453 AAAATCAAGCATACAACTAGTGTATAAAGTTCCAAATCAACAAATACAGGTAA 2512  
 Db 425 hrAsnGlnAlaGlnGluProGlyVallyLysLeu-----ArgGlnAlaenLysValV 443  
 QY 2513 TCAGTAAAGAAATTTCTGAATTAATGACCGTGAATATGAAAGATTTAAACCATCAAGATT 2572  
 Db 443 alLysLysArgAsnSerLysLeuSerAlaIleGlyPheGluGluPheGluAspGlnAspT 463  
 QY 2573 GCGAAGGAAT 2583  
 Db 463 tpGluGlyAsn 466  
 RESULT 13  
 ABR42346  
 ID ABR42346 standard; protein; 466 AA.  
 AC ABR42346;  
 XX  
 XX 11-AUG-2003 (first entry)  
 XX  
 DE Mouse interphotoreceptor matrix IPM 150, isoform D.  
 XX  
 KW Mouse; interphotoreceptor matrix; IPM 150; IPMC; receptor;  
 KW ophthalmological; gene therapy.  
 XX  
 OS Mus sp.

XX Key Location/Qualifiers  
 FH Peptide 1..20  
 FT /label= Signal\_peptide 21..466  
 FT Protein /label= IPM\_150 21..49  
 FT Domain /note= "conserved domain" 360..408  
 FT Domain /note= "EGF-like domain"  
 XX WO2003039346-A2.  
 PN 15-MAY-2003.  
 PD 08-NOV-2002; 2002WO-US036090.  
 PF 08-NOV-2001; 2001US-00077270.  
 PR (IOWA ) UNIV IOWA RES FOUND.  
 PA Hageman GS, Kuehn MH;  
 XX WPI; 2003-441440/41.  
 DR N-PSDB; AC057951.  
 XX  
 PT New interphotoreceptor matrix proteins and polynucleotides, useful for  
 PT treating or preventing photoreceptor death or retinal detachment, or for  
 PT treating ocular disorders.  
 XX Claim 6; Page 83; 105pp; English.  
 CC The present sequence is the protein sequence of isoform D of novel mouse  
 CC interphotoreceptor matrix 150 (IPM 150), a member of the newly identified  
 CC interphotoreceptor matrix component (IPMC) gene family. IPMC gene family  
 CC members have been identified in humans, monkey, cow, goat, rabbit, dog,  
 CC cat, pig, mouse and rat IPM. 2 Subfamilies are designated IPM 150 (or  
 CC IPMG1) and IPM 200 (or IPMG2). The invention provides IPM 150 and IPM 200  
 CC polynucleotides and polypeptides, antibodies that specifically bind the  
 CC polypeptides, and vectors comprising the polynucleotides. A claimed  
 CC method of treating or preventing photoreceptor death or retinal  
 CC detachment involves administering an IPMC polynucleotide, polypeptide or  
 CC antibody. Also claimed is a method for identifying a compound capable of  
 CC modulating IPMC gene expression  
 XX SQ Sequence 466 AA;  
 Alignment Scores:  
 Pred. No.: 7.73e-111 Length: 466  
 Score: 1298.50 Matches: 309  
 Percent Similarity: 43.76% Conservative: 52  
 Best Local Similarity: 37.45% Mismatches: 99  
 Query Match: 22.10% Indels: 365  
 DB: 7 Gaps: 10  
 US-10-007-270-1 (1-3330) x ABR42346 (1-466)  
 QY 131 ATGTATTTTGGAAACTAGAAAGTATTTTGTGTTTGGATTTTCTCCAGTCAAGA 190  
 Db 1 MetAsnPheGlnIleLysHisAlaIlePheValPheGlyIlePheLeuGlnValGlnGly 20  
 QY 191 ACCAAGATATCTCCATTAACATATACCATCTCTGAAACTAAGACATAGACAAATCCCCA 250  
 Db 20 ----- 20  
 QY 251 AGAATGAACAACATGAAGTACTGAAAAATCTACAAAATGTCACTATGAGACGAATA 310  
 Db 20 ----- 20  
 QY 311 TTCGATTTGGCAAGCATCGAACAAAAAGATCGCATTTTCCCAACGGGGTTAAAGTC 370  
 Db 20 ----- 20

|    |      |  |  |  |  |  |      |
|----|------|--|--|--|--|--|------|
| Db | 337  | heValGluProA <sup>g</sup> LeuGluAlaValAsp-                     |  |  |  |  | 346  |
| QY | 1493 | GGTCTCCACCTGCTATGAGGCTCTACCTCCCTGTGTCAGAGCTCCACCTTCTTTATGGCAT  |  |  |  |  | 1552 |
| Db | 346  |  |  |  |  |  | 346  |
| QY | 1553 | CAAGCATCTTCTCTGTGACTGATCAAGGCACCAAGATACAATGCCACTGCACAGCAA      |  |  |  |  | 1612 |
| Db | 346  |  |  |  |  |  | 346  |
| QY | 1613 | TGCTAGTACAGGGCTCACCATCCCCACAGTGAATTATTCGCAATCAGCCAACTGGGTC     |  |  |  |  | 1672 |
| Db | 346  |  |  |  |  |  | 346  |
| QY | 1673 | TGGGAATTTCAATCCACCTGCATCTTCAGATGACAGCCGATCAAGTGACGGTGGCGAAG    |  |  |  |  | 1732 |
| Db | 346  |  |  |  |  |  | 346  |
| QY | 1733 | ATATGGTCAGACACCTAGATGAATGGATCTGTCTGACACTCTCGCCCATCTGAGGTAC     |  |  |  |  | 1792 |
| Db | 346  |  |  |  |  |  | 346  |
| QY | 1793 | CAGAGCTCAGGGAATATGTTTCTGTCTCCAGATCATTTCTTGGAGGATPACCATCTCTGTCT |  |  |  |  | 1852 |
| Db | 346  |  |  |  |  |  | 346  |
| QY | 1853 | CAGCTTTACAGTATATCACCACCTAGTTCTATGACCATTTGCCCCAGAGCCGCGAGCTGG   |  |  |  |  | 1912 |
| Db | 346  |  |  |  |  |  | 346  |
| QY | 1913 | TAGTGTCTTCACTCTGGTGTGTTCTTAACATGGCTTCTCCAGACCTGTTTCAACAAGA     |  |  |  |  | 1972 |
| Db | 346  |  |  |  |  |  | 346  |
| QY | 1973 | GCTCTCTGGAGTACCGAGCTCTGGAGCAACAATTCACACAGCTGCTGGTTCCTCATCTAC   |  |  |  |  | 2032 |
| Db | 346  |  |  |  |  |  | 346  |
| QY | 2033 | GATCCAACTCTTACAGATTTAAGCAACTTGAATACTTAACTTCAGAAACGGAGTGTGA     |  |  |  |  | 2092 |
| Db | 347  |  |  |  |  |  | 352  |
| QY | 2093 | TTGTGAATAGCAAAATGAAGTTTGTAAAGTCTGTGGCTATTAACCTCACAAGGCTGCG     |  |  |  |  | 2152 |
| Db | 352  | eu-  |  |  |  |  | 352  |
| QY | 2153 | ACGGGCTCTGGAGGATTTGTTCTGCTGACGCCAACAACCTCATCTGGAAATAGACA       |  |  |  |  | 2212 |
| Db | 352  |  |  |  |  |  | 352  |
| QY | 2213 | GCTACTCTCTCAACATGATCAACAGCTGATCAAGCAGATCCCTGCAAGTTCTCTGGCTGCG  |  |  |  |  | 2272 |
| Db | 353  |  |  |  |  |  | 353  |
| QY | 2273 | GGGAATTTGCCAATGTGTAAAGAACGACGACTGAGGAAGCGAGGTCTCGCTGCAAAAC     |  |  |  |  | 2332 |
| Db | 366  | lYlYsPheAlaGlnCysVallyAsnGluTrpThrGluGluAlaGluCysArgCysArgG    |  |  |  |  | 386  |
| QY | 2333 | CAGGATATGACACCGAGGCGCTGGACGGTCTGGAACACAGGCTCTGTGGCCCTGGCA      |  |  |  |  | 2392 |
| Db | 386  | InglyHlsgLusHlsgLysLeuAspYrGlnThrLeuAsnLeuCysProGly-           |  |  |  |  | 405  |
| QY | 2393 | CAAGGAATGCGAGGTCTCTCAGGGAAGGAGCTCCATGCAAGTGGCCAGATCACTCTG      |  |  |  |  | 2452 |
| Db | 406  | --lyeThrCysValAlaGlyArgGluGlnAlaThrProCysArgProThrAspHiserT    |  |  |  |  | 425  |
| QY | 2453 | AAATCAAGCATACAAAACCTAGTGTGTAAAGTTCAAAATCAACAAAATAACAGTAA       |  |  |  |  | 2512 |
| Db | 425  | hrAsnGlnAlaGlnProGlyVallyAsyLeu-----ArgGlnAsnlyValy            |  |  |  |  | 443  |
| QY | 2513 | TCAGTAAAGAAATTCGAATTTCTGACCTGTAGATATATGAGAAATTAACCATCAAGAT     |  |  |  |  | 2572 |

Db 443 aliylsYsrghsnSerlysluSerAlaalleGlyPheGluGluPheGluAspGlnAspT 463  
QY 2573 GGGAGCAAAAT 2583  
DB: |||||  
463 rpGluGlyAsn 466  
RESULT 14  
ADAI4845  
ID ADAI4845 standard; protein; 198 AA.  
AC ADAI4845;  
XX  
DT 06-NOV-2003 (first entry)  
XX  
DE Human interphotoreceptor matrix component, IPMC, 150 isoform C.  
XX  
KW human; IPMC 150 isoform C; gene therapy;  
KW interphotoreceptor matrix component; IPMC; ocular disorder;  
KW macular degeneration; photoreceptor death; retinal detachment.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..20  
FT Protein 21..198 /note= "Signal sequence"  
FT /note= "Mature IPMC 150 isoform C"  
FT Misc-difference 39 /note= "Encoded by NCC"  
FT Misc-difference 66 /note= "Encoded by NAT"  
XX  
PN US2002160954-A1.  
XX  
PD 31-OCT-2002.  
XX  
PF 08-NOV-2001; 2001US-00007270.  
XX  
PR 29-OCT-1998; 98US-00183972.  
XX  
PR 29-OCT-1999; 99US-00430195.  
XX  
PA (IOWA ) UNIV IOWA RES FOUND.  
XX  
PI Hageman GS, Kuehn MH;  
XX  
DR WPI; 2003-238235/23.  
DR N-PSDB; ADAI4844.  
XX  
PT New isolated or recombinant interphotoreceptor matrix component  
PT polynucleotide and polypeptide, useful for diagnosing, preventing,  
PT treating or prognosticating ocular disorders, e.g. macular degeneration  
PT or retinal detachment.  
XX  
PS Claim 8; Page 37-38; 76pp; English.  
XX  
CC The invention relates to an isolated or recombinant interphotoreceptor  
CC matrix component (IPMC) polynucleotide. Also disclosed is a vector  
CC comprising a promoter of an interphotoreceptor matrix component (IPMC)  
CC gene operatively linked to the IPMC polynucleotide. The IPMC  
CC polynucleotides, polypeptides and antibodies are useful for diagnosing,  
CC preventing, treating or prognosticating ocular disorders, e.g. macular  
CC degeneration, photoreceptor death or retinal detachment. They are also  
CC useful for identifying a compound capable of modulating IPMC gene  
CC expression in a cell. The present sequence represents the amino acid  
CC sequence of human interphotoreceptor matrix component, IPMC, 150 isoform  
CC C.  
SQ Sequence 198 AA;  
Alignment Scores:  
Pred. No.: 6.82e-87 Length: 198  
Score: 1038.00 Matches: 197  
Percent Similarity: 99.49% Conservative: 0

Best Local Similarity: 99.49% Mismatches: 1  
Query Match: 17.67% Indels: 0  
DB: 6 Gaps: 0  
US-10-007-270-1 (1-3330) x ADAI4845 (1-198)  
QY 131 ATGTATTGGAACTAGAGAGCTATTATTTTGGATTCTTCCAGTTCACAGGA 190  
DB 1 MetTyrLeuGluThrArgArgAlaIlePheValPheThrPilePheLeuGlnValGIngly 20  
QY 191 ACCAAGATATCTCCATTAAATATATACCATCTGAAATTAAGACATAGACATCCCCCA 250  
DB 21 ThrLysAspIleSerIleAsnIleTyrHisSerGluThrLysAspIleAsnProPro 40  
QY 251 AGAAATGAACAACTGAAAGTACTGAAATATGACAAATGCAACTATGAGACGAAAT 310  
DB 41 ArgAsnGluThrGluSerThrGluLysMetTyrLysMetSerThrMetArgArgile 60  
QY 311 TTCGATTGGCAAGCATCGAACAAAGATCCGATTTTCCCAACGGGGTTAAAGTC 370  
DB 61 PheAspLeuAlaLysHisArgThrLysArgSerAlaPhePheProThrGlyValLysVal 80  
QY 371 TGTCACAGCAATCCATGAAACAGATTTAGACAGTCTTCAAGCTTATTATAGATTGAGA 430  
DB 81 CysProGlnGluSerMetLysGlnIleLeuAspSerLeuGlnAlaTyr-TyrArgLeuArg 100  
QY 431 GTGTGTGAGGAGCAGTATGGGAAGCATATCGGATCTTCTGGATCGCATCCCTGACACA 490  
DB 101 ValCysGlnGluAlaIleTrpGluAlaTyrArgIlePheLeuAspArgileProAspThr 120  
QY 491 GGGGAATATCAGGACTGGGTGACGATCTGCCAGCAGGAGACCTTCTGCTCTTTGACATT 550  
DB 121 GlyGluTyrGlnAspTrpValSerIleCysGlnGlnGluThrPheCysLeuPheAspIle 140  
QY 551 GGAAAAAACTTCAGCAATTCACAGGAGCACTGATCTTCCACAGACAGAGATAAAGACAG 610  
DB 141 GlyLysAsnPheSerAsnSerGlnGluHisLeuAspLeuGlnArgIleLysGln 160  
QY 611 AGAAGTTTCCCTCGACAGAAAGATGAAATATCTGCAGAGACAGACATTGGGAGAGCTGGT 670  
DB 161 ArgSerPheProAspArgLysAspGluIleSerAlaGluLysThrLeuGlyGluProGly 180  
QY 671 GAAACATTGTCTTTCACAGCAATCTACATTTTCAAGACTTGGCAGATATTC 724  
DB 181 GluThrIleValIleSerThrAlaIleTyrIleSerLysThrTrpAlaValPhe 198  
RESULT 15  
ABR42344  
ID ABR42344 standard; protein; 198 AA.  
XX AC ABR42344;  
XX  
DT 11-AUG-2003 (first entry)  
XX  
DE Human interphotoreceptor matrix IPM 150, isoform C.  
XX  
KW Human; interphotoreceptor matrix; IPM 150; IPMC; receptor;  
KW ophthalmological; gene therapy.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..20 /label= "Signal\_peptide"  
FT Protein 21..198 /label= "IPM 150"  
FT Misc-difference 39 /note= "encoded by NCC"  
FT Misc-difference 66 /note= "encoded by NAT"  
FT Domain 71..78 /note= "N-terminal domain"  
FT Domain 95..115

/note= "conserved domain"

FT XX WO2003039346-A2.  
 XX PN 15-MAY-2003.  
 XX PD 08-NOV-2002; 2002WO-US036090.  
 XX PF 08-NOV-2001; 2001US-00077270.  
 XX PR (IOWA ) UNIV IOWA RES FOUND.  
 XX PA Hageman GS, Kuehn MH;  
 XX PI WPI; 2003-441440/41.  
 XX DR N-PSDB; ACC57948.  
 XX RX New interphotoreceptor matrix proteins and polynucleotides, useful for  
 PT treating or preventing photoreceptor death or retinal detachment, or for  
 PT treating ocular disorders.  
 XX PS Claim 6; Page 80; 105pp; English.  
 XX CC The present sequence is the protein sequence of isoform C of novel human  
 CC interphotoreceptor matrix 150 (IPM 150), a member of the newly identified  
 CC interphotoreceptor matrix component (IPMC) gene family. The IPM 150 gene  
 CC is located on chromosome 6q13-q15, a region that also contains loci for  
 CC progressive bifocal choriorretinal atrophy, autosomal dominant Stargard's  
 CC -like macular dystrophy, North Carolina macular dystrophy and Salla  
 CC disease. Members of the IPMC gene family have been identified in humans,  
 CC monkey, cow, goat, rabbit, dog, cat, pig, mouse and rat IPM. 2  
 CC Subfamilies are designated IPM 150 (or IPM1) and IPM 200 (or IPM2). The  
 CC invention provides IPM 150 and IPM 200 polynucleotides and polypeptides,  
 CC antibodies that specifically bind the polypeptides, and vectors  
 CC comprising the polynucleotides. A claimed method of treating or  
 CC preventing photoreceptor death or retinal detachment involves  
 CC administering an IPMC polynucleotide, polypeptide or antibody. Also  
 CC claimed is a method for identifying a compound capable of modulating IPMC  
 CC gene expression  
 XX SQ Sequence 198 AA;

Search completed: March 4, 2004, 18:45:30  
 Job time : 198.5 secs

Alignment Scores:  
 Pred. NO.: 6.82e-87 Length: 198  
 Score: 1038.00 Matches: 197  
 Percent Similarity: 99.49% Conservative: 0  
 Best Local Similarity: 99.49% Mismatches: 1  
 Query Match: 17.67% Indels: 0  
 DB: 7 Gaps: 0

US-10-007-270-1 (1-3330) x ABR42344 (1-198)

QY 131 ATGATTTTGGAACTAGAAGCTATTTTGGTATTTTGGATTTTCCCAAGTCAAGGA 190  
 Db 1 MetTyrLeuGluThrArgAlaIlePheValPheTyrIlePheLeuGlnValGlnGly 20  
 QY 191 ACCAAGATATCTCCATTACATATACCATCTGAACTTAAAGACATAGACAATCCCCCA 250  
 Db 21 ThrLysAspIleSerIleAsnIleTyrHisSerGluThrLysAspIleAspAsnProPro 40  
 QY 251 AGAAATGAACAACACTGAAGTACTGAAAATCTGCAAAATCTCAACTATGAGACGAATA 310  
 Db 41 ArgAsnGluThrThrGluSerThrGluLysMetTyrLysMetSerThrMetArgIle 60  
 QY 311 TTCGATTTGGCAAGCATCGAACAAGATCGGCAATTTTCCCAACGGGGTTAAAGTC 370  
 Db 61 PheAspLeuAlaLysHisArgThrLysArgSerAlaPhePheProThrGlyValVal 80  
 QY 371 TGTCACAGGAATCCATGAACAGATTTTAGACAGTCTTCAAGCTTATATATAGATTGAGA 430  
 Db 81 CysProGlnGluSerMetLysGlnIleLeuAspSerLeuGlnAlaTyrTyrArgLeuArg 100  
 QY 431 GTGTGTGCAGGAAGCAGTATGCGGAGCATATCCGATCTTCTGGATCGCATCCCTGCACAA 490

Db 101 ValCysGlnGluAlaAlaTyrGluAlaTyrArgIlePheLeuAspArgIleProAspThr 120  
 QY 491 GGGGATATCAGGACTGGGTTCAGCATCTGCCAGAGAGACCTTCTGCTCTTTGACATT 550  
 Db 121 GlyGluTyrGlnAspTyrValSerIleCysGlnGlnGluThrPheCysLeuPheAspIle 140  
 QY 551 GGAATAAACTTCCAGCAATTTCCAGGAGACCTGCTCTCCAGCAGAGAAATAAACAG 610  
 Db 141 GlyLysAsnPheSerAsnSerGlnGluHisLeuAspLeuGlnGlnArgIleLysGln 160  
 QY 611 AGAAGTTTCCCTGCACAGAAAAGATGAAATATCTGCAGAGACATTTGGGAGAGCCTGGT 670  
 Db 161 ArgSerPheProAspArgLysAspGluIleSerAlaGluLysThrLeuGlyGluProGly 180  
 QY 671 GAAACCATTTGTCATTTCAACAGCAATCTACATTTCAAGACTTGGGAGATTTTC 724  
 Db 181 GluThrIleValIleSerThrAlaIleTyrIleSerLysThrTrpAlaValPhe 198

| Result No. | Score | Query |        | DB | ID                  | Description         |
|------------|-------|-------|--------|----|---------------------|---------------------|
|            |       | Match | Length |    |                     |                     |
| 1          | 148.5 | 2.5   | 947    | 4  | US-09-418-780A-1    | Sequence 1, Appli   |
| 2          | 148.5 | 2.5   | 947    | 4  | US-09-392-714-23    | Sequence 23, Appl   |
| 3          | 146.5 | 2.5   | 2137   | 4  | US-09-134-001C-4463 | Sequence 4463, Appl |
| 4          | 146.5 | 2.5   | 2409   | 6  | 5160808-2           | Patent No. 5160808  |
| 5          | 145.5 | 2.5   | 2035   | 2  | US-08-479-537A-2    | Sequence 2, Appli   |
| 6          | 145.5 | 2.5   | 2035   | 3  | US-09-083-116-2     | Sequence 2, Appli   |
| 7          | 145.5 | 2.5   | 2035   | 4  | US-09-134-916A-2    | Sequence 2, Appli   |
| 8          | 140.5 | 2.4   | 529    | 1  | US-08-178-477B-32   | Sequence 32, Appl   |
| 9          | 140.5 | 2.4   | 529    | 4  | US-09-304-121-2     | Sequence 2, Appli   |
| 10         | 140.5 | 2.4   | 783    | 4  | US-09-513-783A-176  | Sequence 176, App   |
| 11         | 139   | 2.4   | 1481   | 2  | US-08-616-844-40    | Sequence 40, Appl   |
| 12         | 139   | 2.4   | 1481   | 2  | US-08-599-654-40    | Sequence 40, Appl   |

|    |      |   |      |
|----|------|---|------|
| QY | 224  | GAAACTAAGACATGACAATACTCCGCCAACAAATCAAACTGAAGAATGCTAAAAAATG                                | 283  |
| Db | 158  | SerAlalysGluLysSerSerPro<br>     <br>:::<br>:::<br>:::                                    | 171  |
| QY | 284  | TACAAATGTCAACTATGACAGCAGCATATTTCGATTGGCAAAGCATCGAACAAGAATCC                               | 343  |
| Db | 172  | PhelyeGlnGlnIleProSerValPheProLysThr-Serile<br>:::<br>:::<br>:::<br>:::                   | 186  |
| QY | 344  | GCATTTTCCCACAGGGGGTTAAAGTCTGTCTCCACAGCAATCCATGAAAACAGATTTTAGAC                            | 403  |
| Db | 187  | -----SerProLeuAsn-<br>:::<br>:::<br>:::   | 190  |
| QY | 404  | AGTCTCAAGCTTATTATAGATTGAGAGTGTCGACGAAGCAGTATGGGACGATATCGG                                 | 463  |
| Db | 191  | -----ValValGlnGlyAlaSerValAsnSerSe<br>   <br>   <br>                                      | 200  |
| QY | 464  | ATCTTTCTCGATCGCATCCCTGACACAGGGGAATATCAGG-   | 503  |
| Db | 200  | rSerGlnThrAlaGlnValThrLysGlyValLysArgLysAlaAspThrThrThrPr<br>   <br>   <br>               | 220  |
| QY | 504  | ---ACTGGGTCAACATCTGCCACGACGACCTTCCTGCTTTCACATTCGGAAGAAAC                                  | 559  |
| Db | 220  | oAlaThrSerAlaValLysAlaSerGluPheSerProThrPheThr---GluLysSe<br>:::<br>:::<br>:::<br>:::     | 239  |
| QY | 560  | TTCAGCAATT-----CCAGGAGCACCTGGATCTCTCCAGCAG  | 598  |
| Db | 239  | rValAlaLeuProProlLysGluAsnMetProLysAsnValLeuProAspSerGlnGl<br>   <br>   <br>              | 259  |
| QY | 599  | AGAAATAAAC----AGAGAAGTTTCCCTGCACGAAAGATCAATATCTGCAGAGAA-                                  | 651  |
| Db | 259  | nGlnTyrsenValValGluThrValLysValThrGluGlnLeuArgHis-Cyese-GluI<br>:::<br>:::<br>:::<br>:::  | 279  |
| QY | 652  | -----GACATTTGGGAGAGCCTCG-   | 669  |
| Db | 279  | leLeuLysGluMetLeuAlaLysLysHisPheSerTyrrAlaTrpProPheTyrsenProv<br>   <br>   <br>           | 299  |
| QY | 670  | -----TGAAACCATTTGCTATTCCAACGCAATCTACATTTCAAAGACT  | 712  |
| Db | 299  | alAspValAsnAlaLeuGlyLeuHisAsnTy-TyrAspValVallLysAsnProMetAspL<br>:::<br>:::<br>:::<br>::: | 319  |
| QY | 713  | TGGGCAAGTATTCTAAGAAA  | 732  |
| Db | 319  | eUGlyThrIleLysGluLysMetAspAsnGlnLutLyrsAspAlaTy-Ser-PheAlaa<br>   <br>   <br>             | 339  |
| QY | 733  | -----CCCTCAGAGAG--  | 744  |
| Db | 339  | laaspValArgLeuMetPheMetAsnCysTyrlLysTyrsenProProAspHisGluVal<br>   <br>   <br>            | 359  |
| QY | 745  | -----CAAATTCAAGTGTGCCAAGCTCTCAGTGGGCTTCCCTCTCA  | 790  |
| Db | 359  | alThrMetAlaArgMetLeuGlnAspValPheGluThrHisPheSerLysIleProIleG<br>:::<br>:::<br>:::<br>:::  | 379  |
| QY | 791  | CTCCTGTGATGACACC-----CTCCTCAATGAATTTCTCGAATAATCACTCA                                      | 835  |
| Db | 379  | luProvalGluSerMetProLeuCytTrilLysThrAspIleThrGluThrThrGli<br>   <br>   <br>               | 399  |
| QY | 836  | ACGACACCAAGATPGCCTACACAGAAAGAGAAACAGAAATTCGCTGTGTGGAGGACAGA                               | 895  |
| Db | 399  | rgGluasnThrAsnGlnAlaSerSerGluGlyAsenSerSerAspAspSerGluaspGlu<br>:::<br>:::<br>:::<br>:::  | 419  |
| QY | 896  | GGGTGGAGCTACGCTCTCTCTGGTAAACCAAGATTTCAAGGACAGCTCGTGACTCC                                  | 955  |
| Db | 419  | rgVal-----LysArgLeuAlaLysLeuGlnGlnL<br>   <br>   <br>                                     | 430  |
| QY | 956  | AGTCCCCATATTACCAGGAGCTAGCAGGAAGATCCCACTTCAGATCGTCTGACTCC                                  | 1015 |
| Db | 430  | eUlysAlaValHisGlnGlnLeuGlnValLeuSerGlnValProPheArgLysLeuAsnL<br>:::<br>:::<br>:::<br>:::  | 450  |
| QY | 1016 | AGAACTTCCAGGATTCAAAAAATCCATGTGTAGGATTTAGACCAAGCAAGAAAAAG                                  | 1075 |
| Db | 450  | yLysLysGluLysSerLysLys-----GluLysLysGluLysv<br>   <br>   <br>                             | 464  |
| QY | 1076 | ATGGCTCAAGCTCC-----ACAGATGCAACTTACGGCCATCT  | 1114 |

|    |      |  |                              |
|----|------|--|------------------------------|
| Db | 464  | alAsnAsnSerAsnGluAsnProArgLysMetCysGluGlnMetArgLysLeuLysGluLysS 484  | :::    ::: :::    :::    ::: |
| QY | 1115 | TPAGAGACAGCTGCAGAGCAAGCAAAAGCCCTCCAAAGTGAACCTCTCTCTCTTTGATTCCA 1174  | ::: :::                      |
| Db | 484  | erLysArgAsn-----GlnProLysLysArgLysGlnGlnPheIleGly----- 498           | :::                          |
| QY | 1175 | ACAAATATGAAGGTGAGGAAGTCTATCATGGACACATGGAGGAGACAGACACACAGAA 1234      | :::                          |
| Db | 499  | -----LeuLysSerGluAsp-----GluAspAsnAlaLysProMetA 511                  | :::                          |
| QY | 1235 | TCTATCTCACAGCTACAGACCTCAAAAGCGTGATCAGCAAAAGCACACTAGAGGAGAAACAAT 1294 | :::                          |
| Db | 511  | bnfYr-----AspGluLysArgGlnLeuSerLeuAsnIleAsnLysLeuProG 527            | :::                          |
| QY | 1295 | CTTTGGATGTGGGACAAATTCAGTTCTACTCATGCAAAATTGCTGGATCACTGCCCGCTTTG 1354  | :::                          |
| Db | 527  | lyAspLysLeuGlyArgVal-----ValHisIleIleGlnSerArgGluProSerLeuS 545      | :::                          |
| QY | 1355 | GTCTGTGACCAACCAATCAGAGCTGCCACATCTTTTGTTGTTATACAGAGGAGATGCTACTT 1414  | :::                          |
| Db | 545  | erAsnSerAsnProAspGluIleGluLeuAspPheGluThrLeu---LysAlaSerThrL 564     | :::                          |
| QY | 1415 | TGAGTCGCAGAACTCTCTCTGTGACCCGACGCTTGACAGGTGAGCGAGCAGAGCATG 1474       | :::                          |
| Db | 564  | euArg-----GluLeuGluLysTyrValSerAla-----C 574                         | :::                          |
| QY | 1475 | GTCTACCTGCACACTTCTTGGTCTCCACCTGCT-----ATGGCCTCTACCTCCC 1522          | :::                          |
| Db | 574  | ysLeuArgLysArgProLeuLysProAlaLysLysIleMetMetSerLysGluGluL 594        | :::                          |
| QY | 1523 | TGTCAGAGCTCCACTTCTTTTANGGCAATCAAGCATCTTCTCTGTGATCATCAAGGCA 1582      | :::                          |
| Db | 594  | euHisSerGlnLysLysGlnGluLeuGlyArgLeuLeuAspValAsnAsnGlnLeuA 614        | :::                          |
| QY | 1583 | CCACAGATACAAATGGCCACTGACACAGCAATCTAGTACCAGGGCTCACCATCCCC-----A 1639  | :::                          |
| Db | 614  | snSerArgLysArgGlnThrLysSerAspLysThrGlnProSerLysAlaValGluAsnV 634     | :::                          |
| QY | 1640 | CAAGTATATCTGCATACAGCAACTGGCTCTGGGAATTTCAATCCACTGCATCTT 1699          | :::                          |
| Db | 634  | alSerArgLeuSerGluSerSerSerSerSerSerSerGluSerGluSerSerS 654           | :::                          |
| QY | 1700 | CAGATGAC-----AGCGATCAAGTCAGCTGGCGAGATATGTCAGACACTAG 1750             | :::                          |
| Db | 654  | erSerAspLeuSerSerSerSerSerSerSerSerGluSerGluMetPheProLysPheT 674     | :::                          |
| QY | 1751 | ATGAATATGGATGTGTGCACACTCTGCGCCCATCTGAGGTACACAGCTCAGCGAATATG 1810     | :::                          |
| Db | 674  | hrGluValLysProAsnAspSerProSerLysGluHisValLysLysMetLysAsnGluC 694     | :::                          |
| QY | 1811 | TTTCTGTCCCAAT-----CATTTCTTGGAGGATPACCA 1843                          | :::                          |
| Db | 694  | ysIleLeuProGluGluValArgThrGlyValThrGlnIleGlyTyrCysValGlnAspThrT 714  | :::                          |
| QY | 1844 | CTCTGTCTC---TCAGCTTTACAGTATACACACTAGTCTATGACCATTCGCCCCAGG 1900       | :::                          |
| Db | 714  | hrSerAlaasnThrThrLeuValHisGlnThrThrProSerHisValMetProProAsnH 734     | :::                          |
| QY | 1901 | CCGAGAGCTGTGTAGTGTCTTCAGTCTGCGTGTGTCTAACTAGCGCTTCTCCACAGACC 1960     | :::                          |
| Db | 734  | ishisGlnLeuAla----- 738  | :::                          |
| QY | 1961 | TGTTTCAACAGAGCTCTCTGGATACCGAGCTCTGGAGCAACAATTCA-----CAGC 2014        | :::                          |
| Db | 739  | --PheAsnTyrGlnGluLeuGluHisLeuGlnThrValLysAsnIleSerProLeuGlnI 758     | :::                          |
| QY | 2015 | TGCTGTTCTCCA-----TATCTAGATCCCAATCTTACAGGATTT----- 2052               | :::                          |
| Db | 758  | leLeuProProSerGlyAspSerGluGlnLeuSerAsnGlyIleThrValMetHisProS 778     | :::                          |
| QY | 2053 | -----AAGCAAC 2059  | :::                          |

179 CAAGTTCAGGACCAACAGAT-----ATCCATTACATATACCATCT 223  
     |||||     |||     |||     |||     |||  
 138 GlnValValGlyValIleLeuLysGlyThrGlnGlnAsnIlealaValSer 157  
               |||     |||     |||     |||     |||  
 224 GAACCTAAGCATTAGACAATCCCCAAGAATGAACAACTGAAGACTCGAAAATG 293  
         |||||     |||     |||     |||     |||  
 158 SerAlaLysGluIysSerSerPro-----SerAlaThrGluLysVal 171  
                   |||     |||     |||     |||     |||

|    |      |  |      |
|----|------|--|------|
| Db | 484  | erLysArgAsn-----GlnProLysLysArgLysGlnGlnPheIleLys-----       | 498  |
| QY | 1175 | ACAAATTTGAAGTGTGAGGAAGTCTATCATGACCAATGAGGAGGACAAGCAACCAAGAA  | 1234 |
| Db | 499  | -----LeuLysSerGluAsp-----GluAspAsnAlaLysProMetA              | 511  |
| QY | 1235 | TCTATCTCACAGCTACAGACCTCAAAAGGCTGATCAGCAAGACACTAGTAGGAGAACAA  | 1294 |
| Db | 511  | snfYr-----AspGluLysArgGlnLeuSerLeuAsnIleAsnLysLeuProG        | 527  |
| QY | 1295 | CTTTGGATGTGGGACAAATTCAGTTCATCATGAATTCCTGGATCTATGCCAGCCTTTG   | 1354 |
| Db | 527  | lyAspLysLeuGlyArgVal-----ValHisIleIleGlnSerArgGluProSerLeuS  | 545  |
| QY | 1355 | GTCTCGACACCCCAATCAGAGCTGCCCATCTTTTGTGTTATACAGAGGATGCTACTT    | 1414 |
| Db | 545  | erAsnSerAsnProAspGluIleGluIleAspPheGluThrLeu---LysAlaSerThrL | 564  |
| QY | 1415 | TGAGTCACAGACTCTCTCTGTGTGAACCCAGCTTGAGACAGTGGACGGAGCAGAGCATG  | 1474 |
| Db | 564  | euArg-----GluLeuGluLysTyrValSerAla-----C                     | 574  |
| QY | 1475 | GTCTACCTGACACTTCTGGTCTCCACTGCT-----ATGGCTCTTACCTCCC          | 1522 |
| Db | 574  | ysLeuArgLysArgProLeuLysProAlaLysIleMetMetSerLysGluGluL       | 594  |
| QY | 1523 | TGTCAAGAGCTCCACCTTCTTTATGTCATCAAGCATCTCTCTCTGATGATCAAGGCA    | 1582 |
| Db | 594  | euHisSerGlnLysLysGlnGluLeuGluLysArgLeuLeuAspValAsnAsnGlnLeuA | 614  |
| QY | 1593 | CCACAGATACATGGCCTCTGACAGACATCTAGTACAGGGCTCCACATCCCC---A      | 1639 |
| Db | 614  | snSerArgLysArgGlnThrLysSerAspLysThrGlnProSerLysAlaValGluAsnV | 634  |
| QY | 1640 | CCAGTGATTTCTGCAATCAGCCAACTGGCTCTGGGAATTCACATCCATGCATCTT      | 1699 |
| Db | 634  | alSerArgLysSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerS | 654  |
| QY | 1700 | CAGATGAC-----AGCCGATCAAGTCGAGGTGGCGAAGATATGGTCAGACACCTAG     | 1750 |
| Db | 654  | erSerAspLeuSerSerSerAspSerAspSerAspSerGluSerGluMetPheProLysP | 674  |
| QY | 1751 | ATGNAATGATCTGTGACACTCTGCCCCCATCTGAGGTACAGAGCTCAGCGAATATG     | 1810 |
| Db | 674  | hrGluValLysProAsnAspSerProSerLysGluHisValLysLysMetLysAsnGluC | 694  |
| QY | 1811 | TTTCTGTGCCAGAT-----CAATCTTGAGAGATACCA                        | 1843 |
| Db | 694  | ysIleLeuProGluLysArgThrGlyValThrGlnIleGlyTyrCysValGlnAspThrT | 714  |
| QY | 1844 | CTCTGTC---TCAGCTTTACAGTATATCACCATAGTCTTATGACCATCCGCCCAAGG    | 1900 |
| Db | 714  | hrSerAlaAsnThrThrLeuValHisGlnThrThrProSerHisValMetProAsnH    | 734  |
| QY | 1901 | GCCGAGAGCTGTAGTGTCTTCAGTCTGCTGTGCTAACATGGCCCTTCCCAACGACC     | 1960 |
| Db | 734  | ishiGlnLeuAla-----   | 738  |
| QY | 1961 | TGTTTCAACAGAGCTCTCTGAGTACCGAGCTCTGGAGCAACAATTACA---CAGC      | 2014 |
| Db | 739  | --PheAsnTyrGlnGluLeuGluHisLeuGlnThrValLysAsnIleSerProLeuGlnI | 758  |
| QY | 2015 | TGCTGTTTCCA-----TATCTACGATCCAATCTTACAGATT---                 | 2052 |
| Db | 758  | leLeuProProSerGlyAspSerGluGlnLeuSerAsnGlyIleThrValMetHisProS | 778  |
| QY | 2053 | -----AAGCAAC   | 2059 |
| Db | 778  | erGlyAspSerAspThrThrMetLeuGluSerGluCysGlnAlaProValGlnLysAspI | 798  |
| QY | 2060 | TTCAAAATCTTAATCTTCAGAACCGGAGTGTGATTGTGAATAGCAAAATGAAATTGCTA  | 2119 |

798 leYsIlleYsAen-----AlaAspSerTrpYsSerLeuGlyL 811

2120 AGCTCTGTG---CCGTATAACCTCACCAAGCTGTGCACGGGTCITGGAGGATTTCCGTT 2176

811 ysProValYsProSerGlyValMetYsSerSerAspGluLeuPheAenGlnPheArg 831

2177 CTGTCTGCA-----GCCCAACAACTC-----CATCTCG 2203

831 ysAlaAlaIleGluYsGluValYsAlaAargThrGlnGluLeuIleArgYsHisLeuG 851

2204 AAATAGACAGCTACTCTCTCAACATTGAACACAGCTGATCAAGCAGATCCCTCGCAAGTCC 2263

851 luGlnAenThrYsGluLeuYsAlaSerGlnGluAenGlnAargAspLeuGlyAenGlyL 871

2264 TGGCTCTGGCGGAATTGCCCAATGTGTAAAGAACGAGACTGAGGAAGCGAGTGTCT 2323

871 euThrValGluSerPheSerAsnYsIleGlnAsnYsCysSerGlyGluGluGln-Lys 890

2324 GCTCAAAACACGAGATATGACAGCAGCGAGCGCTTGGACGCTCTGCAACAGGCTCTGTG 2383

891 GluHisProGlnSerSerGluIleGlnIleAspYsSerYsLeuTrpLeuLeuYsAspArg 910

2384 GCCCTGGCAACAAGAAATGGAGTCTCTCCAGGAAGAGGAGCTCCATCGAGGT 2437

911 AspleuAlaArgProYsGluGlnGluIleArgArgGluAlaMetValGly 928

RESULT 3

US-09-134-001C-4463

; Sequence 4463, Application US/09134001C

; Patent NO. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 4463

; LENGTH: 2137

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4463

|                        |          |               |                     |          |  |
|------------------------|----------|---------------|---------------------|----------|--|
| Alignment Scores:      |          |               |                     |          |  |
| Pred. No.:             | 0.000752 | Length:       | 2137                |          |  |
| Score:                 | 146.50   | Matches:      | 124                 |          |  |
| Percent Similarity:    | 35.23%   | Conservative: | 86                  |          |  |
| Best Local Similarity: | 20.11%   | Mismatches:   | 298                 |          |  |
| Query Match:           | 2.49%    | Indels:       | 89                  |          |  |
| DB:                    | 4        | Gaps:         | 15                  |          |  |
| US-10-007-270-1        | (1-3330) | x             | US-09-134-001C-4463 | (1-2137) |  |

|    |      |   |      |
|----|------|---|------|
| QY | 221  | TCTGAACACTAAGACATACAGCAANTCCCCCGAAGAATGAAACAACACTGAAAGTACTGAAAA | 280  |
|    |      | :::     :::     :::     :::     :::     :::     :::     :::     |      |
| Db | 1104 | SerAspSerAlaSerThr-SerThrSerGlusSerAspSerThrSerGlusSerThrSerLeu | 1123 |
| QY | 281  | ATGTCAAAGATGTCACACTATGAGACGAATAATTCGATTTTGGCAGACGATCGACAAAAGA   | 340  |
|    |      | :::     :::     :::     :::     :::     :::     :::     :::     |      |
| Db | 1124 | SerGlusSerLeuSerThr-----SerValSerAspSerThrSerAlaSerThrSerGlu    | 1141 |
| QY | 341  | TCGCATTTTTCCCAACGGGGTTAA-----                                   | 366  |
|    |      | :::     :::     :::     :::     :::     :::     :::             |      |
| Db | 1142 | SerAlaSerThrSerThrSerGlusSerGlusSerAsnSerAlaSerThrSerLeuSerGly  | 1161 |
| QY | 367  | AGTCGTGCCACAGGAATCATGTAA-----ACAGATTTTAGACAGCTCTCAAGCTTA        | 417  |
|    |      | :::     :::     :::     :::     :::     :::     :::             |      |
| Db | 1162 | SerLeuSerThrSerIleSerAspSerThrSerThrSerThrSerAspSerAlaSerThr    | 1181 |



```

QY 418 TTATAGATTGAGAGTGTCTCAGGAAGCAGATATGGAAGCATATCGATCTTTCTGGATCG 477
Db 1182 SerThrSerGluSerGluSerThrSerThrSerLeuSerGluSerThrSerThr 1201
QY 478 CATCCCTGACACAGGGGAATATCAGAGCTGGGTGAGCATCTCCAGCAGAGACCTTCGTG 537
Db 1202 --SerLeuSerAspSerThrSerThrSerGluSerAlaSerThrSerThrSerG 1221
QY 538 CTTCTTTGACATTGGAAAAAATCTTCAGCAATTCAGGAGACCTGGATCTTCTCCAGCA 597
Db 1221 luSerAspSerThrSerGluSerThrSerLeuSerGluSerThr 1235
QY 598 GAGATAAAACAGAGAAGTTTCCCTGCACAGAAAAGATGAATATCTGCAGAGAAGACATT 657
Db 1236 -----SerThrSerValSerAspSerThrSerAlaSerThrSerAspSerAlaSerT 1253
QY 658 GGGAGAGCTCGCTGAAACCAATGTCTTCAACAGCAATCTACATTTCAAGACTTGGGC 717
Db 1253 hrSerThrSerValSerAspSerGluSerAlaSerThrSerIleSerGlu-SerLeuSer 1272
QY 718 AGTATTCTAAGAAAACCTCAGAGAGCAAAATTCAGATGTGGCAAGCTCTCACTTGGG 777
Db 1273 ThrSerValSerAspSerThrSerThrSerSerAspSerAlaSerThrSerThrSer 1292
QY 778 CTTTCCCTCTCTACTCTGATGACACCTCTCTCAATGAAATTTCTGATATACACTCAAC 837
Db 1293 GluSerAspSerThrSerGluSerThrSerLeuSerGluSerIleSerThrSerValSer 1312
QY 838 GACACCAAGATGCTCAACAGAGAAAGAAACAGAAATTCGCTGTGGTGGAGGAGCAGAGG 897
Db 1313 AspSerThrSerAlaSerThrSerAspSerAlaSerThrSerThrSerGluSerGluSer 1332
QY 898 GTGGAGCTCAGCCTCTCTCTGTTAAACAGAGTTTCAAGGGCAGAGCTCGCTCACTCCAG 957
Db 1333 AspSerAlaSerThrSerLeu-SerGlySerThrSerThrSerLeuSerAspSerThr 1351
QY 958 TCCCATATTACAGAGAGTACGA-----GGAAAGTCCCAACTTCAGATGCAAAAGATA 1011
Db 1352 SerThrSerThrSerAspSerAlaSerThrSerThrSerGluSerAspSerGluSerGlu 1371
QY 1012 TTAAAGAAATCTCCAGATTCAAAAATCCATGTGTAGGATTTAGCAAAAGAAAGAA 1071
Db 1372 SerThrSerLeuSerGlySerThrSerThrSerLeuSerAspSerThrSerThrSer 1391
QY 1072 AAAGATGGCTCAAGCTCCACAGAGATCGCAACTTACGCCATCTTTAAGAGACAGAGTCA 1131
Db 1392 SerAspSerAlaSerThrSer-----ThrSerValSerAspSerAsnSerAla 1407
QY 1132 GAAGCA-----AAAAGCCCTGCAAGTGACCTCTGCTTTTGTATTCCAAACAAATTCGA 1185
Db 1408 SerThrSerLeuSerGlySerLeuSerThrSerValSerAspSerThrSerThrSer 1427
QY 1186 AGTCGAGAGTCTATCATGGAAACCATGGAGGAGGAGCAAGCAACCAAAATCTATCTACA 1245
Db 1428 SerAspSerAlaSerThrSerThrSerGluSerAspSerGlu-----ArgAla 1443
QY 1246 GCTACAGACCTCAAGGCTGATCAGCAAGCATCTAGAGGAAGAAACAATCTTTGGATGTG 1305
Db 1444 SerThrSerLeuSerGlySerThrSerThrSerIleSerAspSerThrSerThrSer 1463
QY 1306 GGGCAATTCAGTTCATCGATGAATGCTGGATCATCGCCACCTTTGGTCTTCGACACC 1365
Db 1464 SerAspSerAlaSerThrSerThrSerValSerGluSerAsnSerThrSerThrSerIle 1483
QY 1366 CAATCAGAGCTGCCACATCTTTGTGTTAATACAGAGGATGCTACT-----1413
Db 1484 SerGluSerLeuSerThrSerValSerAspSerThrSerThrSerSerAspSerAla 1503
QY 1414 -----TTGAGTCCAGAACTTCTCTGTTGAACCCAGCTTGACAGAGTG 1458
Db 1504 SerThrSerThrSerValSerAspSerSerAlaSerThrSerSerSerGluSerVal 1523

```

```

QY 1459 GACGGAGCAGAGCATGGTCTTACTGACATCTTCTGGTCTCCACCTCTATGGCCTTACC 1518
Db 1524 SerThrSerAspSerGluSerThrSerThrSerThrSerAspSerAlaSerThrSerThr 1543
QY 1519 TCCCTGTCAAGAGCTCCACCTTTCTTTATGGCANTCAGCATCTTCTCTCTGACTGATCAA 1578
Db 1544 SerValSerGlu-----SerAsnSerThrSerThrSerLeuSer 1556
QY 1579 GGCACACAGATACAAATGGCCACTGCACAGACAATGCTAGTACCAGGGCTCACCATCCCC 1638
Db 1557 GlySerThrSerThrSerValSerAspSerThr-----SerThrSer 1570
QY 1639 ACCAGTGAATTATTCGAATCAGCCAACTGGCTCTGGGAATTTCACTCCACTGCACTCT 1698
Db 1571 ThrSerAspSerAlaSerAlaSer-----ThrSerGluSer 1582
QY 1699 TCAGATGACAGCCCATCAAGTCAGGTGGCGAAGATATGTCAGACACCTAGATGAATG 1758
Db 1583 AspSerAspSerAlaSerThrSerSerSerGluSerVal-----SerThr 1597
QY 1759 GATCTCTGACACTCTCTGCCCATCTGAGGTACAGGTACAGAGCTCAGCGAATATGTTCTGTC 1818
Db 1598 SerValSerAspSerThrSerAlaSerThrSerGluSerAlaSerThrSerThrSerVal 1617
QY 1819 CCAGATCAT-----TTCCTTGGAGGAT 1839
Db 1618 SerAspSerAsnSerAlaSerThrSerLeuSerGluSerThrSerThrSerLeuSerAsp 1637
QY 1840 ACCACTCTCTGCTCAGCTTTTACAGTATATACCACTAGTTCT 1881
Db 1638 SerThrSerMetSerThrSerAspSerAlaSerThrSerThr 1651

RESULT 4
5180808-2
; Patent No. 5180808
; APPLICANT: RUOSLAHTI, ERKKI I.
; TITLE OF INVENTION: VERSION CORE PROTEIN NUCLEIC ACID
; SEQUENCES ENCODING THE SAME, NUCLEIC ACID PROBES, ANTI-VERSION
; ANTIBODIES, AND METHODS OF DETECTING THE SAME
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/441,179
; FILING DATE: 27-NOV-1989
; SEQ ID NO:2:
; LENGTH: 2409
5180808-2

Alignment Scores:
Pred. No.: 0.00803 Length: 2409
Score: 146.50 Matches: 184
Percent Similarity: 29.66% Conservative: 115
Best Local Similarity: 18.25% Mismatches: 313
Query Match: 2.49% Indels: 396
DB: 6 Gaps: 48

US-10-007-270-1 (1-3330) x 5180808-2 (1-2409)
QY 691 ACCAATCTACATTTCAAGACTTGGCAGTATTCTTAAGAAAACCTCAGAGAGCAATT 750
Db 1468 SerSerThrThrPheValSerAspGlySerLeuGluLysHisPro-----GluVal 1484
QY 751 CAAGATGTTCCAAAGCTCTCACTTGGCCCTTCCCTCTCACT-----792
Db 1485 ProSerAlaLysAlaValThrAlaAspGlyPheProThrValSerValMetLeuProLeu 1504
QY 793 -----CCTGATGACACCTCTCTCAAT-----813
Db 1505 HisSerGluGlnAsnLysSerSerProAspProThrSerThrLeuSerAsnThrValSer 1524
QY 814 -----GAAATTCGATAATACA 831
Db 1525 TyrGluArgSerThrAspGlySerPheGlnAspArgPheArgGluPheGluAspSerThr 1544

```



Db 2183 AspTyrGlyTrpHisLysPheGlnGlyGlnCysTyrLysTyrPheAlaHisArgArgThr 2202  
QY 2440 CCAGATCACTCTGAATAATCAA-----GCATACAAACTAGTGTAAAG 2484  
Db 2203 TrpAspAlaAlaGluArgGluCysArgLeuGlnGlyAlaHisLeuThrSerIleLeuSer 2222  
QY 2485 TTCCAAATCAACA-----AAT 2502  
Db 2223 HisGluGlnGlnMetPheValAsnArgValGlyHisAspTyrGlnTrpIleGlyLeuAsn 2242  
QY 2503 AACAGGTAAATCAGTAA-----AGAAATCTGAATTAAGTACGCTAGATATGAGAA 2556  
Db 2243 AspLysMetPheGluHisAspPheArgTrpThrAspGlySerThrLeuGlnTyrGluAsn 2262  
QY 2557 TTT-----AACCATCAAGAT----- 2571  
Db 2263 TrpArgProAsnGlnProAspSerPheSerAlaGlyGluAspCysValValIleIle 2282  
QY 2572 TGGGAAGGAATTAATAAATCGAA-----ATGTACAAATTAATCACTTAGGCTAT---CTC 2622  
Db 2283 TrpHisGluAsnGlyGlnTrpAsnAspValProCysAsnTyrHisLeuThrTyrThrCys 2302  
QY 2623 AGAGAGATGATTG-----CCTTCTCAAGAAATGGAGACAGGCATAT 2667  
Db 2303 LysLysGlyThrValAlaCysGlyGlnProProValValGluAsnAla-----LysTh 2320  
QY 2668 TCATGGTCATCAAAATCCAGACATACAGTCAACACTGAGAAATCAGCACACACCATATTT 2727  
Db 2320 rPheGlyLysMetLysProArgTyrGluIleAsnSer----- 2332  
QY 2728 CAAATATAGAGAGTCAATGATCTGCGAACCACTAAATCTGAAAAAAGACACTTACT 2787  
Db 2333 -----LeuIleArgTyrHisCysLysAspGlyPhe-----IleGlnArgHisLeuPr 2348  
QY 2788 TATTATTAA----- 2802  
Db 2348 oThrIleArgCysLeuGlyAsnGlyArgTyrAlaIleProLysIleThrCysMetAsnPr 2368  
QY 2803 AAATGCAATCAGCAACATAT 2824  
Db 2368 oSerAlaTyrGlnArgThrTyr 2375  
RESULT 5  
US-08-479-537A-2  
; Sequence 2, Application US/08479537A  
; Patent No 5861381  
; GENERAL INFORMATION:  
; APPLICANT: CHAMBER, Pierre  
; APPLICANT: KIENY, Marie-Paule  
; APPLICANT: LATHE, Richard  
; APPLICANT: HAREUVENI, Mara  
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE  
; TREATMENT OR PREVENTION OF A MALIGNANT TUMOR  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/479,537A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 90/13101  
; FILING DATE: 23-OCT-1990

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: WO PCT/FR91/00835  
;; FILING DATE: 23-OCT-1991  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/039,320  
;; FILING DATE: 04-APR-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/403,576  
;; FILING DATE: 14-MAR-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Teskin, Robin L.  
;; REGISTRATION NUMBER: 35,030  
;; REFERENCE/DOCKET NUMBER: 017753-025  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703) 836-6620  
;; TELEFAX: (703) 836-2021  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2035 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FEATURE:  
;; NAME/KEY: Peptide  
;; LOCATION: 128..1899  
;; OTHER INFORMATION: /note= "The amino acids spanning  
;; OTHER INFORMATION: 128 to 1899 constitute a repeated  
;; OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such  
;; OTHER INFORMATION: repeats varies from 1 to 40."  
;; FEATURE:  
;; NAME/KEY: Peptide  
;; LOCATION: 134  
;; OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa  
;; OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,  
;; OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."  
;; FEATURE:  
;; NAME/KEY: Peptide  
;; LOCATION: 144  
;; OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa  
;; OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA,  
;; OTHER INFORMATION: or ACG; and Asn = AAT or AAC."  
;; FEATURE:  
;; NAME/KEY: Peptide  
;; LOCATION: 147  
;; OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa  
;; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA,  
;; OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."  
;; FEATURE:  
;; NAME/KEY: Peptide  
;; LOCATION: 1..21  
;; OTHER INFORMATION: /note= "Amino acids 1 to 21 are a  
;; OTHER INFORMATION: 21 amino acid precursor sequence."  
;; US-08-479-537A-2  
Alignment Scores:  
Pred. No.: 0.00897 Length: 2035  
Score: 145.50 Matches: 74  
Percent Similarity: 38.11% Conservative: 51  
Best Local Similarity: 22.56% Mismatches: 144  
Query Match: 2.48% Indels: 59  
DB: 2 Gaps: 13  
US-10-007-270-1 (1-3330) x US-08-479-537A-2 (1-2035)  
QY 1357 CTGACACCAATCAGAGTCCCATCTTTGTTGTTATACAGAGGATCTACTTTG 1416  
Db 1622 ProAsp\*\*\*Arg---Pro\*\*\*ProGlySerThrAlaPro\*\*\*AlaHisGlyValThrSer 1640  
QY 1417 AGTCAGAACTTCTCCT-----GTTGACCCAGCTTGCAGACGTGACGGA 1464  
Db 1641 AlaProAsp\*\*\*ArgPro\*\*\*ProGlySerThrAlaPro\*\*\*AlaHisGlyValThrSer 1660

QY 1465 GCAGAGCATGCTACCTGACACT--TCTTGGTCTCCACCTGCTGTATGGCTCTACCTCC 1521  
Db 1466 GCGAGAGCATGCTACCTGACACT--TCTTGGTCTCCACCTGCTGTATGGCTCTACCTCC 1521  
Db 1467 GCGAGAGCATGCTACCTGACACT--TCTTGGTCTCCACCTGCTGTATGGCTCTACCTCC 1521  
QY 1522 CTGTCCAGAGCTCCACCTTCTTTATGGCATCAAGCATCTCTCTCTGACTGATCAAGGC 1581  
Db 1523 CTGTCCAGAGCTCCACCTTCTTTATGGCATCAAGCATCTCTCTCTGACTGATCAAGGC 1581  
Db 1524 CTGTCCAGAGCTCCACCTTCTTTATGGCATCAAGCATCTCTCTCTGACTGATCAAGGC 1581  
QY 1582 ACCAGAGTACATGGCCACTGACGACATGCTA-----GTACACGGG 1626  
Db 1583 ACCAGAGTACATGGCCACTGACGACATGCTA-----GTACACGGG 1626  
Db 1584 ACCAGAGTACATGGCCACTGACGACATGCTA-----GTACACGGG 1626  
QY 1690 SerThrAlaPro\*\*\*AlaHisGlyValThrSerAlaProAsp\*\*\*ArgPro\*\*\*ProGly 1709  
Db 1691 SerThrAlaPro\*\*\*AlaHisGlyValThrSerAlaProAsp\*\*\*ArgPro\*\*\*ProGly 1709  
Db 1692 SerThrAlaPro\*\*\*AlaHisGlyValThrSerAlaProAsp\*\*\*ArgPro\*\*\*ProGly 1709  
QY 1627 CTCCACATCCCACTGCT-----GATTATCTGCAATCAGCCCACTGCTGCTGGA 1677  
Db 1628 CTCCACATCCCACTGCT-----GATTATCTGCAATCAGCCCACTGCTGCTGGA 1677  
Db 1629 CTCCACATCCCACTGCT-----GATTATCTGCAATCAGCCCACTGCTGCTGGA 1677  
QY 1710 SerThrAlaPro\*\*\*AlaHisGlyValThrSerAlaProAsp\*\*\*ArgPro\*\*\*LeuGly 1729  
Db 1711 SerThrAlaPro\*\*\*AlaHisGlyValThrSerAlaProAsp\*\*\*ArgPro\*\*\*LeuGly 1729  
Db 1712 SerThrAlaPro\*\*\*AlaHisGlyValThrSerAlaProAsp\*\*\*ArgPro\*\*\*LeuGly 1729  
QY 1678 ATTCCATCCCACTGCT-----GCATCTTCAGATGACGCGCATCAAGTCAAGTGGC 1728  
Db 1679 ATTCCATCCCACTGCT-----GCATCTTCAGATGACGCGCATCAAGTCAAGTGGC 1728  
Db 1680 ATTCCATCCCACTGCT-----GCATCTTCAGATGACGCGCATCAAGTCAAGTGGC 1728  
QY 1729 GAAGATATGTCAGACACTGATGATGATCTCTGACACTCTGCTGCTGCTGCTGCTGCTG 1788  
Db 1730 GAAGATATGTCAGACACTGATGATGATCTCTGACACTCTGCTGCTGCTGCTGCTGCTG 1788  
Db 1731 GAAGATATGTCAGACACTGATGATGATCTCTGACACTCTGCTGCTGCTGCTGCTGCTG 1788  
QY 1750 SerThrLeuValHisGlyValThrSerAlaArgAlaThrThrProAlaSerLysSer 1769  
Db 1751 SerThrLeuValHisGlyValThrSerAlaArgAlaThrThrProAlaSerLysSer 1769  
Db 1752 SerThrLeuValHisGlyValThrSerAlaArgAlaThrThrProAlaSerLysSer 1769  
QY 1789 GTACAGAGCTCCAGCAATATGTTCTGCTCCAGATCAATTTCTTGGAGGATCACTCT 1848  
Db 1790 GTACAGAGCTCCAGCAATATGTTCTGCTCCAGATCAATTTCTTGGAGGATCACTCTCT 1848  
Db 1791 GTACAGAGCTCCAGCAATATGTTCTGCTCCAGATCAATTTCTTGGAGGATCACTCTCT 1848  
QY 1770 ThrProPro-----SerLeuProSerHisHisSerAspThrProThrThr 1784  
Db 1771 ThrProPro-----SerLeuProSerHisHisSerAspThrProThrThrThr 1784  
Db 1772 ThrProPro-----SerLeuProSerHisHisSerAspThrProThrThrThr 1784  
QY 1849 GTC-----TCAGCTTTACAGTAT 1866  
Db 1850 GTC-----TCAGCTTTACAGTAT 1866  
Db 1851 GTC-----TCAGCTTTACAGTAT 1866  
QY 1785 LeuAlaSerHisSerThrLysThrAspAlaSerSerThrHisHisSerThrValProPro 1804  
Db 1786 LeuAlaSerHisSerThrLysThrAspAlaSerSerThrHisHisSerThrValProPro 1804  
Db 1787 LeuAlaSerHisSerThrLysThrAspAlaSerSerThrHisHisSerThrValProPro 1804  
QY 1867 ATCAGCATGTTCTATGACATGCTCCCAAGGCGGAGAGCTGTGTCTTCTTC--- 1923  
Db 1868 ATCAGCATGTTCTATGACATGCTCCCAAGGCGGAGAGCTGTGTCTTCTTC--- 1923  
Db 1869 ATCAGCATGTTCTATGACATGCTCCCAAGGCGGAGAGCTGTGTCTTCTTC--- 1923  
QY 1805 LeuThrSerSerHisSerThrLysThrProGlnLeuSerThrGlyValSerPhePhe 1824  
Db 1806 LeuThrSerSerHisSerThrLysThrProGlnLeuSerThrGlyValSerPhePhe 1824  
Db 1807 LeuThrSerSerHisSerThrLysThrProGlnLeuSerThrGlyValSerPhePhe 1824  
QY 1924 ---AGTCTGCTGTTCTAACTGCTCTCTCCAGACCTGTTCTCAACAGAGCTCTG 1980  
Db 1925 ---AGTCTGCTGTTCTAACTGCTCTCTCCAGACCTGTTCTCAACAGAGCTCTG 1980  
Db 1926 ---AGTCTGCTGTTCTAACTGCTCTCTCCAGACCTGTTCTCAACAGAGCTCTG 1980  
QY 1825 LeuSerPheHisLeuSerHisGlnPheAsnSerSerLeuGlnAspProSerThrAsp 1844  
Db 1826 LeuSerPheHisLeuSerHisGlnPheAsnSerSerLeuGlnAspProSerThrAsp 1844  
Db 1827 LeuSerPheHisLeuSerHisGlnPheAsnSerSerLeuGlnAspProSerThrAsp 1844  
QY 1981 GAGTACAGCTGCTGAGCAACAATTCACAGCTGCTGCTGCTTCCATATCTACGATCCAAT 2040  
Db 1982 GAGTACAGCTGCTGAGCAACAATTCACAGCTGCTGCTGCTTCCATATCTACGATCCAAT 2040  
Db 1983 GAGTACAGCTGCTGAGCAACAATTCACAGCTGCTGCTGCTTCCATATCTACGATCCAAT 2040  
QY 1845 TyrTyrGlnGluLeuGlnArgAspIleSerGluMetPheLeuGlnAsnTyrLysGlnGly 1864  
Db 1846 TyrTyrGlnGluLeuGlnArgAspIleSerGluMetPheLeuGlnAsnTyrLysGlnGly 1864  
Db 1847 TyrTyrGlnGluLeuGlnArgAspIleSerGluMetPheLeuGlnAsnTyrLysGlnGly 1864  
QY 2041 CTACAGATTTAAGCACTGAATATTAATCTCAACAGAGCTGTGTGTGTGAAT 2100  
Db 2042 CTACAGATTTAAGCACTGAATATTAATCTCAACAGAGCTGTGTGTGTGAAT 2100  
Db 2043 CTACAGATTTAAGCACTGAATATTAATCTCAACAGAGCTGTGTGTGTGAAT 2100  
QY 1865 -----GlyPheLeuGlyLeuSerAsnIleLysPheArgProGluSerValValGln 1882  
Db 1866 -----GlyPheLeuGlyLeuSerAsnIleLysPheArgProGluSerValValGln 1882  
Db 1867 -----GlyPheLeuGlyLeuSerAsnIleLysPheArgProGluSerValValGln 1882  
QY 2101 AGCAAAATGAAGTTTCTAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2160  
Db 2102 AGCAAAATGAAGTTTCTAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2160  
Db 2103 AGCAAAATGAAGTTTCTAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2160  
QY 1883 LeuThrLeuAlaPheArgGlnGly-----ThrIleAsnValHisAspVal 1897  
Db 1884 LeuThrLeuAlaPheArgGlnGly-----ThrIleAsnValHisAspVal 1897  
Db 1885 LeuThrLeuAlaPheArgGlnGly-----ThrIleAsnValHisAspVal 1897  
QY 2161 -----TTGAGGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2211  
Db 2162 -----TTGAGGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2211  
Db 2163 -----TTGAGGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2211  
QY 1898 GluThrGlnPheAsnGlnTyrLysThrGluAlaAlaSerArgTyrAsnLeuThrIleSer 1917  
Db 1899 GluThrGlnPheAsnGlnTyrLysThrGluAlaAlaSerArgTyrAsnLeuThrIleSer 1917  
Db 1900 GluThrGlnPheAsnGlnTyrLysThrGluAlaAlaSerArgTyrAsnLeuThrIleSer 1917  
QY 2212 AGTACTCTCTCAACATGAACCA 2235  
Db 2213 AGTACTCTCTCAACATGAACCA 2235  
Db 2214 AGTACTCTCTCAACATGAACCA 2235  
QY 1918 AspValSerValSerHisValPro 1925  
Db 1919 AspValSerValSerHisValPro 1925  
Db 1920 AspValSerValSerHisValPro 1925

## RESULT 6

US-09-083-116-2  
; Sequence 2, Application US/09083116  
; Patent No. 6203795  
; GENERAL INFORMATION:  
; APPLICANT: CHAMBERON, Pierre  
; APPLICANT: KIENY, Marie-Paule  
; APPLICANT: LATHE, Richard  
; APPLICANT: HAREUVENI, Mara  
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE  
; TREATMENT OR PREVENTION OF A MALIGNANT TUMOR  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,116  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/479,537  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR91/00835  
FILING DATE: 23-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/039,320  
FILING DATE: 04-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/403,576  
FILING DATE: 14-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teekin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 017753-025  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2035 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 128..1899  
OTHER INFORMATION: /note= "The amino acids spanning 128 to 1899 constitute a repeated region wherein the repeat is 20 amino acids, 17 of which are fixed. The number of such  
OTHER INFORMATION: repeats varies from 1 to 40."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 134  
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa  
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,  
OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 144  
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa  
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA,  
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 147  
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa  
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA,  
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..21  
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a  
OTHER INFORMATION: 21 amino acid precursor sequence."  
US-09-083-116-2

Alignment Scores:

|                        |          |               |      |
|------------------------|----------|---------------|------|
| Pred. No.:             | 0.000897 | Length:       | 2035 |
| Score:                 | 145.50   | Matches:      | 74   |
| Percent Similarity:    | 38.11%   | Conservative: | 51   |
| Best Local Similarity: | 22.56%   | Mismatches:   | 144  |
| Query Match:           | 2.48%    | Indels:       | 59   |
| DB:                    | 3        | Gaps:         | 13   |

US-10-007-270-1 (1-3330) x US-09-083-116-2 (1-20335)

|    |      |  |      |
|----|------|--|------|
| QY | 1357 | CTGTGACACCCAAATCAGAGCTGCCACATCTTTGCTGTATTATAACAGAGGAGTCTACTTTG | 1411 |
| Db | 1622 | ProAsp***Arg---Pro***ProGlySerThrAlaPro***AlaHisGlyValThrSer   | 1640 |
| QY | 1417 | AGTCAGACACTTCCTCT  | 1464 |
| Db | 1641 | AlaProAsp***ArgPro***ProGlySerThrAlaPro***AlaHisGlyValThrSer   | 1660 |
| QY | 1465 | GCAGAGCATGCTTACTCGACACT---TCCTTGCTCCACCTGCTATGGCCCTCTACCTCC    | 1521 |
| Db | 1661 | AlaProAsp***ArgPro***ProGlySerThrAlaPro***AlaHisGlyValThrSer   | 1680 |
| QY | 1522 | CTGTGACAGCTCCACTTCTTTATGGCATCAAGCATCTTCTCTGACTGATCAAGGC        | 1581 |
| Db | 1681 | AlaProAsp***ArgPro---***ProGly                                 | 1689 |
| QY | 1582 | ACCACAGATCAATGCCACTGACCCAGACAATGCTA---GTACCAAGG                | 1626 |
| Db | 1690 | SerThrAlaPro***AlaHisGlyValThrSerAlaProAsp***ArgPro***ProGly   | 1709 |
| QY | 1627 | CTCACCATCCCAAGT---GATTATTCTGCAATCAGCCAACTGGCTCGGA              | 1677 |
| Db | 1710 | SerThrAlaPro***AlaHisGlyValThrSerAlaProAsp***ArgPro***LeuGly   | 1729 |
| QY | 1678 | ATTTCACATCCACT---GCATCTCAGAGACAGCCGATCAAGTCAGAGTGC             | 1728 |
| Db | 1730 | SerThrAlaProProValHisAsnValThrSerAlaSerGlySerAlaSerGlySerAla   | 1749 |
| QY | 1729 | GAAGATATGTCAGACACTAGATGAATGGATCTGTCGACACTCTGCGCCCATCTGAG       | 1788 |
| Db | 1750 | SerThrLeuValHisAsnGlyThrSerAlaAlaThrThrProAlaSerLysSer         | 1769 |
| QY | 1789 | GTACAGAGCTCAGCGAATATGTTCTGTCCAGCATCATTTCTTGAGGATACCACTCT       | 1848 |
| Db | 1770 | ThrProPro---SerLeProSerHisSerAspThrProThrThr                   | 1784 |
| QY | 1849 | GTC---TCAGCTTTACGAT  | 1866 |
| Db | 1785 | LeuAlaSerHisSerThrLysThrAspAlaSerSerThrHisSerThrValProPro      | 1804 |
| QY | 1867 | ATCACCACTAGTTCTATGACCATTCGCCCAAGGGCCGAGCTGGTAGTCTTCTC---       | 1923 |
| Db | 1805 | LeuThrSerSerAsnHisSerThrSerProGlnLeuSerThrGlyValSerPhePhe      | 1824 |
| QY | 1924 | ---AGTCGTGGTTGTCAATGGCTTCTCCACAGCCTGTTCAACAGAGCTCTCTG          | 1980 |
| Db | 1825 | LeuSerPheHisLeuSerAsnLeuGlnPheAsnSerSerLeuGluAspProSerThrAsp   | 1844 |
| QY | 1981 | GAGTACCGAGCTCTGGAGCAACATTCACACAGCTGCTGCTCCATATCTACAGATCCAAT    | 2040 |
| Db | 1845 | TyrTyTGlnGlnLeuGlnArgAspPheLeuGlnAsnTyLysGlnGly                | 1864 |
| QY | 2041 | CTTACAGGATTTAAGCAACTGGAATACTTAATCTTCAGAACGGGAGTGATCTGGAAT      | 2100 |
| Db | 1865 | ---GlyPheLeuGlyLeuSerAsnIleLysPheArgProGluSerValValGln         | 1882 |
| QY | 2101 | AGCAAAATGAGTTTGCTTAAGTCTGCGGTATTAACCTCACCAAGGCTGTGACAGGGTTC    | 2160 |
| Db | 1883 | LeuThrLeuAlaPheArgGluGly---ThrIleAsnValHisAspVal               | 1897 |
| QY | 2161 | ---TTGGAGGATTTTCGTTCTGCTGCACCCCAACAATCTCCATCTGGAAATAGAC        | 2211 |
| Db | 1898 | GluThrGlnPheAsnGlnTyLysThrGluAlaSerArgTyAsnLeuThrIleSer        | 1917 |

|    |      |                          |      |
|----|------|--------------------------|------|
| QY | 2212 | AGCTACTCTCTCAACATTGAACCA | 2235 |
|    |      | : :                      |      |
| Db | 1918 | AspValSerValSerHisValPro | 1925 |

## RESULT 7

US-09-134-916A-2  
Sequence 2, Application US/09134916A  
Patent No. 6328956  
GENERAL INFORMATION:  
APPLICANT: CHAMON, Pierre  
APPLICANT: KIENY, Marie-Paule  
APPLICANT: LATHE, Richard  
APPLICANT: HAREUVENT, Mara  
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE  
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/134,916A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/479,537  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: FR 90/13101  
FILING DATE: 23-OCT-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR91/00835  
FILING DATE: 23-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/039,320  
FILING DATE: 04-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/403,576  
FILING DATE: 14-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 017753-025  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2035 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 128..1899  
OTHER INFORMATION: /note= "The amino acids spanning  
128 to 1899 constitute a repeated region which  
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number  
of repeats varies from 1 to 40."  
OTHER INFORMATION: repeats varies from 1 to 40."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 134  
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa  
Xaa Xaa which is the codon for Pro or Ala which

```

; OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
;
; NAME/KEY: Peptide
; LOCATION: 144
; OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
; OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
; OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
;
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 147
; OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
; OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
;
; NAME/KEY: Peptide
; LOCATION: 1..21
; OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
; OTHER INFORMATION: 21 amino acid precursor sequence."
;
US-09-134-916A-2

Alignment Scores:
Pred. No.: 0.000897 Length: 2035
Score: 145.50 Matches: 74
Percent Similarity: 38.11% Conservative: 51
Best Local Similarity: 22.56% Mismatches: 144
Query Match: 2.48% Indels: 59
DB: 4 Gaps: 13

US-10-007-270-1 (1-3330) x US-09-134-916A-2 (1-2035)
QY 1357 CTGACACCAATCAGAGCTGCCACATCTTTGCTTTATACAGAGGATGCTACTTTG 1416
Db 1622 ProAsp***Arg---Pro***ProGlySerThrAlaPro***AlaHisGlyValThrSer 1640
QY 1417 AGTCAGAACTTCTCTCT-----GTTGAACCCAGCTTGAGAGGAGCGG 1464
Db 1641 AlaProAsp***ArgPro***ProGlySerThrAlaPro***AlaHisGlyValThrSer 1660
QY 1465 GCAGAGCATGCTCTACTGACACT---TCTGTCTCCACCTGTATGGCCTTACCTCC 1521
Db 1661 AlaProAsp***ArgPro***ProGlySerThrAlaPro***AlaHisGlyValThrSer 1680
QY 1522 CTGTCAGAGCTCCACCTTTCTTTATGGCATCAAGCATCTTCTCTGATGATCAAGC 1581
Db 1681 AlaProAsp***ArgPro***ProGlySerThrAlaPro***AlaHisGlyValThrSer 1699
QY 1582 ACCAGATACATGCGCTGACGACGACGACGACGACGACGACGACGACGACGACGAC 1626
Db 1690 SerThrAlaPro***AlaHisGlyValThrSerAlaProAsp***ArgPro***ProGly 1709
QY 1627 CTCACCATCCCAACAGT-----GATTATCTGCACGACGACGACGACGACGACGAC 1677
Db 1710 SerThrAlaPro***AlaHisGlyValThrSerAlaProAsp***ArgPro***LeuGly 1729
QY 1678 ATTTCATCCACCT-----GCATCTTCAGATGACGACGACGACGACGACGACGACGAC 1728
Db 1730 SerThrAlaProValHisAsnValThrSerAlaSerGlySerAlaSerGlySerAla 1749
QY 1729 GAAGATATGTCAGACCTAGATGAATGATCTGCTGACACTCTGCGCCCATCTGAG 1788
Db 1750 SerThrLeuValHisAsnGlyThrSerAlaArgAlaThrThrProAlaSerLysSer 1769
QY 1789 GTACACGAGCTCAGCGAATATGTTCTGCTCCGATCATTTCTTGGAGATACCACTCT 1848
Db 1770 ThrProPro-----SerileProSerHisHisSerAspThrProThrThr 1784
QY 1849 GTC-----TCAGCTTACAGTAT 1866
Db 1785 LeuAlaSerHisSerThrLysThrAspAlaSerSerThrHisHisSerThrValProPro 1804
QY 1867 ATCAGACATGTTCTATGACCATTTGCCCGCCAGGCGGAGAGCTGGTGTCTTC 1923
Db 1805 LeuThrSerSerAsnHisSerThrSerProGlnLeuSerThrGlyValSerPhePhe 1824

```

```

QY 1924 ---AGTCTCGGTGTGCTTAACATGCTTCCAAAGACCTGTTCAACAAGAGCTCTCTG 1980
Db 1825 LeuSerPheHisIleSerAsnLeuGlnPheAsnSerSerLeuGluAspProSerThrAsp 1844
QY 1981 GAGTACGAGCTCTGGAGCAACAATTCACAGAGCTGCTGTTCCATATCTACGATCAAT 2040
Db 1845 TyrTyrGlnGluLeuGlnArgAspIleSerGluMetPheLeuGlnAsnTyrLysGlnGly 1864
QY 2041 CTTACAGGATTTAAGCAACTTGAATATTAACTTCAGAAACGGAGCTGCTGTTGTAAT 2100
Db 1865 ---GlyPheLeuGlyLeuSerAsnIleLysPheArgProGluSerValValGln 1882
QY 2101 AGCAAAATGAAGCTTTGCTTAAGTCTGTGCGGTATTAACCTACCAAGGCTGTGCGGGTTC 2160
Db 1883 LeuThrLeuAlaPheArgGluGly-----ThrIleAsnValHisAspVal 1897
QY 2161 -----TTGGAGGATTTTCTGCTCTGCTCAGCCCAACAACCTCATCTGGAATAAGAC 2211
Db 1898 GluThrGlnPheAsnGlnTyrLysThrGluAlaSerArgTyrAsnLeuThrIleSer 1917
QY 2212 AGCTACTCTCTCAACATTGAACCA 2235
Db 1918 AspValSerValSerHisValPro 1925

RESULT 8
US-08-178-477B-32
; Sequence 32, Application US/08178477B
; Patent No. 5756343
; GENERAL INFORMATION:
; APPLICANT: WU, CARL; CLOS, JOACHIM;
; APPLICANT: WESTWOOD, J. TIMOTHY.; RABINDRAN, SRIDHAR
; TITLE OF INVENTION: CELL STRESS
; TITLE OF INVENTION: TRANSCRIPTION FACTORS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/178,477B
; FILING DATE: 07-JAN-1994
; CLASSIFICATION: 530
; APPLICATION NUMBER: US/07/617,910
; FILING DATE: 26-NOV-1990
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: CAROL M. GRUPPI
; REGISTRATION NUMBER: 37,341
; REFERENCE/DOCKET NUMBER: 2026-4103US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 529
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-178-477B-32

```

Alignment Scores:

Pred. No.: 0.00118  
 Score: 140.50  
 Percent Similarity: 34.24%  
 Best Local Similarity: 21.60%  
 Query Match: 2.39%  
 DB: 1  
 Gaps: 26

US-10-007-270-1 (1-3330) x US-08-178-477B-32 (1-529)

QY 700 CATTTCAAA-----GACTTGGGCGAGTATTCTAAGAAAAACCTCCAGAGACAAATTCAA 753  
 Db 60 TyrPheHisAsnMetAlaSerPheValArgGlnLeuAsnMetTyrGlyPheArg 79  
 QY 754 GATGTTGCCAAGCTCTCACTGGGCTTTCCTCTCTCTCTCT-----GATGACACCTCTC 807  
 Db 80 LysValValHisIleGlnGlnGlyGlyLeu---ValLysProGluArgAspThrGlu 98  
 QY 808 CTC-----AATGAATCTCGAATAATACATCAACGAC 840  
 Db 99 PheGlnHisProCysPheLeuArgGlyGlnGlnLeuGluAsnIleLysArgLys 118  
 QY 841 ACCAAGATCCCTACACAGAAAGAGAAACAGATTCGCTGTGTGGAGAGACAGAGGTG 900  
 Db 119 ValThrSerValSerThrLeuLysSerGluAspIleLysIleArgGlnAspSerValThr 138  
 QY 901 GAGCTCAGCGTCTCTCGTAAACAGAGTTCAAGCAGAGCTCGCTGACTCCAC--- 957  
 Db 139 LysLeuLeuThrAspValGlnLeuMetLysGlyGlnGlnCysMetAspSerLysLeu 158  
 QY 958 -----TCCCATATTACAGAGCTAGCAGGAAGTCCCA--- 993  
 Db 159 LeuAlaMetLysHisGlnAsnGlnAlaLeuTrpArgGluValAlaSerLeuArgGlnLys 178  
 QY 994 ---CTTCAGATCAAAAGATATTAAAGAACTTCCAGGATTC-----AAA 1035  
 Db 179 HisAlaGlnGlnGlnLysValValAsnLysLeuIleGlnPheLeuIleSerLeuValGln 198  
 QY 1036 AAAATCCATGTTGGATTAGACCAAG-----AAAGAAAAGATGGCTCA 1083  
 Db 199 SerAsnArgIleLeuGlyValLysArgLysIleProLeuMetLeuAsnAspSerGlySer 218  
 QY 1084 AGCTCCACAGAGATGCAACTTACGGCCATCTTTAAGACACACAGTGCAGAGCAAAAGC 1143  
 Db 219 AlaHisSerMetProLysTyrSerArgGlnPheSerLeuGluHisValHisGlySerGly 238  
 QY 1144 CTGCAAGTACCTCTCTCTTTGATTCACCAAAATTTGAAGTGAAGTCAATCTATCAT 1203  
 Db 239 Pro-----TyrSerAlaProSerProAlaTyrSerSerSerLeuTyr--- 253  
 QY 1204 GGAACCATGGAGGAGGACAGCAACAGAAATCTATCTCAGAGTACAGACTCAAAAG 1263  
 Db 254 -----AlaProAspAlaValAlaSerSerGlyProfile----- 264  
 QY 1264 CTGATCAGAAAGCACTAGAGAAACAATCTTTGGATGTGGGCAATCTCAGTTCAC 1323  
 Db 265 -----IleSerAspIle 268  
 QY 1324 GATCAAAATCTGGATCACTCCAGCGCTTTGGTCTCT---GACACCAATCAGAGTGCCT 1380  
 Db 269 ThrGluLeuAlaProAlaSerProMetAlaSerProGlyGlySerIleAspLysPro 288  
 QY 1381 ACATCTTTCTGTTTAAACAGAGATGCTACTTTTGATCCAGACTTCTCTCTTTGAA 1440  
 Db 289 LeuSerSerProLeuValArg-----ValLysGluGlnProProSerPro 304  
 QY 1441 CCCAGCTTGAGACAGTGGAGGAGAGAGATGGTCTTACCT-----GACACTTCT 1491  
 Db 305 ProGlnSerProArgValGluGluAlaSerProGlyArgProSerSerValAspThrLeu 324  
 QY 1492 TGGTCTCCACTGCTATGCGCTCTTACCTCCCTGTGACAGAGTCCACTTTCTTTTGGCA 1551  
 Db 325 LeuSerProThrAlaLeuIleAspSerIleLeuArgGlnSerGluPro-----AlaPro 342

QY 1552 TCAAGCATCTCTCTCTGACTGAT---CAAGGACACACAGATACATG----- 1596  
 Db 343 AlaSerValThrAlaLeuThrAspAlaArgGlyHisThrAspThrGluGlyArgProPro 362  
 QY 1596 ----- 1596  
 Db 363 SerProProThrSerThrProGluLysCysLeuSerValAlaCysLeuAspLysAsn 382  
 QY 1597 -----GCCACTGAC-----CAGACATGCTA 1617  
 Db 383 GluLeuSerAspHisLeuAspAlaMetAspSerAsnLeuAspAsnLeuGlnThrMetLeu 402  
 QY 1617 ----- 1617  
 Db 403 SerSerHisGlyPheSerValAspThrSerAlaLeuLeuAspLeuPheSerProSerVal 422  
 QY 1618 ---CTACAGGCTCACCATCCCAAGTATTTCTGCAATCAGC---CAACTGGCT 1671  
 Db 423 ThrValProAspMetSerLeuProAspLeuAspSerSerLeuAlaSerIleGlnGluLeu 442  
 QY 1672 CTGGGAATTTCACTCCACCTGATCTTCAGATGACAGCCGATCAAGTGCA---GGTGGC 1728  
 Db 443 LeuSerProGlnGluProProArgProGluAlaGluAsnSerSerProAspSerGly 462  
 QY 1729 GAAGATATGTCAGACAC-----CTAGAT-----GAA 1755  
 Db 463 LysGlnLeuValHisTyrThrAlaGlnProLeuPheLeuLeuAspProGlySerValAsp 482  
 QY 1756 ATGAGTCTCTGACACTCTCTGCCCTCTGAGGTACAGAGTCCAGGAATATGTTCT 1815  
 Db 483 ThrGlySerAsnAspLeuProValLeuPheGluLeuGlyGlu---GlySerTyrPheSer 501  
 QY 1816 GTCCAGATCATTTCTGGAGGATACCACTCTCTCTCAGCTTTACAGTATATCACCCT 1875  
 Db 502 GluGlyAspGlyPheAlaGluAspProThr-----IleSerLeuLeuThrGly 517  
 QY 1876 AGTTCTATGACATTGCCCCAGGCGGAGAGTGGTAGTG 1917  
 Db 518 SerGlu-----ProProLysAlaLysAspProThrVal 528

RESULT 9

US-09-304-121-2  
 ; Sequence 2, Application US/09304121  
 ; Patent No. 6342596  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Voellmy, Richard  
 ; TITLE OF INVENTION: MOLECULAR REGULATORY CIRCUITS TO ACHIEVE  
 ; FILE REFERENCE: 870109.409  
 ; CURRENT APPLICATION NUMBER: US/09/304,121  
 ; NUMBER OF SEQ ID NOS: 3  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 2  
 ; LENGTH: 529  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapien  
 ; US-09-304-121-2

Alignment Scores:  
 Pred. No.: 0.00118 Length: 529  
 Score: 140.50 Matches: 111  
 Percent Similarity: 34.24% Conservative: 65  
 Best Local Similarity: 21.60% Mismatches: 185  
 Query Match: 2.39% Indels: 153  
 DB: 4 Gaps: 26

US-10-007-270-1 (1-3330) x US-09-304-121-2 (1-529)

QY 700 CATTTCAAA-----GACTTGGGCGAGTATTCTAAGAAAAACCTCCAGAGACAAATTCAA 753  
 Db 60 TyrPheHisAsnMetAlaSerPheValArgGlnLeuAsnMetTyrGlyPheArg 79

|    |      |   |                                   |     |
|----|------|---|-----------------------------------|-----|
| QY | 754  | GATGTTGGCAAGCTCTCACTTGGGCGCTTTCCTCTCACTCT                     | -----GATGACACACCTC                | 807 |
| Db | 80   | LyseValValHisIleGluGlnGlyGlyLeu                               | ---VallysProGluArgAspAspThrGlu    | 98  |
| QY | 808  | CTC   | -----AATGAATTCTCGATAATCACTCAACGAC | 840 |
| Db | 99   | PheGlnHisProCysPheLeuArgGlyGlnGluGlnLeuLeuGluAsnIleLeuArgIlys | 118                               |     |
| QY | 841  | ACCAAGATGCCTACACAGAAAGAAACAGAAATTCGCTGTTGGAGGACAGAGGGT        | 900                               |     |
| Db | 119  | ValThrSerValSerThrLeuLysSerGluAspIleLysIleArgGlnAspSerValThr  | 138                               |     |
| QY | 901  | GAGCTCAGCGTCTCTCGTAAACACAGAGTTCAAGGCAGAGCTCGCTGACTCCCGAG      | 957                               |     |
| Db | 139  | LysLeuLeuThrAspValGlnLeuMetLysGlyLysGlnGluCysMetAspSerLysLeu  | 158                               |     |
| QY | 958  | -----TCCCATATACCAAGAGCTAGCAGGAAGTCCCAA                        | 993                               |     |
| Db | 159  | LeuAlaMetLysHisGluAsnGluAlaLeuThrArgGluValAlaSerLeuArgGlnIys  | 178                               |     |
| QY | 994  | ---CTTCAGATGCAGAAAGATATTTAAGAACTTCACAGATT                     | 1035                              |     |
| Db | 179  | HisAlaGlnGlnIlysValValAsnLysLeuIleGlnPheLeuIleSerLeuValGln    | 198                               |     |
| QY | 1036 | AAATCCATGTGTTAGATTAGACCAAG                                    | 1083                              |     |
| Db | 199  | SerAsnArgIleLeuGlyValLysArgIysIleProLeuMetLeuAsnAspSerGlySer  | 218                               |     |
| QY | 1084 | AGTCCACAGAGATGCACCTTACGGCCNTCTTAAGACACAGTGCAGAGCAAGCAAAAGC    | 1143                              |     |
| Db | 219  | AlaHisSerMetProLysTyrSerArgGlnPheSerLeuGluHisValHisGlySerGly  | 238                               |     |
| QY | 1144 | CCTGCAAGTGACCTCTGCTGTTTGAATCCACAAAAATTCGAAGTGAGGAAGTCTATCAT   | 1203                              |     |
| Db | 239  | Pro-----TyrSerAlaProSerProAlaTyrSerSerSerSerLeuTyr            | 253                               |     |
| QY | 1204 | GGAACCATGAGGAGACACAGCAACACAGAAATCTATCTACAGCTACAGCACTCAAAAGG   | 1263                              |     |
| Db | 254  | -----AlaProAspAlaValAlaSerSerGlyProIle                        | 264                               |     |
| QY | 1264 | CTGATCAGCAACCACTAGAGAGAACCAATCTTGAGTGGGACCAATTCAGTCTACT       | 1323                              |     |
| Db | 265  | -----IleSerAspIle   | 268                               |     |
| QY | 1324 | GATGAATTTCTGGATCACTCCAGCGCTTTGGTCT                            | 1380                              |     |
| Db | 269  | ThrGluLeuAlaProAlaSerProMetAlaSerProGlySerIleAspGluArgPro     | 288                               |     |
| QY | 1381 | ACATCTTTCTGTTATACAGAGGATGCTATTGAGTCCAGAACTTCCTCTCTTGA         | 1440                              |     |
| Db | 289  | LeuSerSerProLeuValArg-----VallysGluGluProProSerPro            | 304                               |     |
| QY | 1441 | CCCAGCTTCAGACAGTCGACGAGCAGAGCATGGTCTACT                       | 1491                              |     |
| Db | 305  | ProGlnSerProArgValGluGluAlaSerProGlyArgProSerSerValAspThrLeu  | 324                               |     |
| QY | 1492 | TGGTCTCCACTGCTATGGCTCTACTCCCTGTCAGAGCTCCCACTTTCTTTATGGCA      | 1551                              |     |
| Db | 325  | LeuSerProThrAlaLeuIleAspSerIleLeuArgGluSerGluPro-----AlaPro   | 342                               |     |
| QY | 1552 | TCAGCATCTTCTCTGACTGAT---CAAGGCACACAGATACAATG                  | 1596                              |     |
| Db | 343  | AlaSerValThrAlaLeuThrAspAlaArgGlyHisThrAspThrGluGlyArgProPro  | 362                               |     |
| QY | 1596 | -----   | 1596                              |     |
| Db | 363  | SerProProThrSerThrProGluLysCysLeuSerValAlaCysLeuAspLysAsn     | 382                               |     |
| QY | 1597 | -----GCCACTGAC  | 1617                              |     |
| Db | 383  | GluLeuSerAspHisLeuAspAlaMetAspSerAsnLeuAspAsnLeuGlnThrMetLeu  | 402                               |     |
| QY | 1617 | -----   | 1617                              |     |

Db 403 SerSerHisGlyPheSerValAspThrSerAlaLeuLeuAspLeuPheSerProSerVal 422  
 QY 1618 ---GTACACAGGGCTCACCATCCACCACAGTGATTATCTGCAATCAGC---CAACTGGCT 1671  
 Db 423 ThrValProAspMetSerLeuProAspLeuAspSerSerLeuAlaSerIleGlnLeu 442  
 QY 1672 CTGGGAATTCACATCCACTGTCATCTTCAGATGACAGCCGATCAAGTGCA---GSTGGC 1728  
 Db 443 LeuSerProGlnGluProProArgProGluAlaGluAsnSerSerProAspSerGly 462  
 QY 1729 GAAGATATGCTCAGACAC-----CTAGAT-----GAA 1755  
 Db 463 IysGlnLeuValHisTyrThrAlaGlnProLeuPheLeuLeuAspProGlySerValAsp 482  
 QY 1756 ATGGATCTGTCTGACATCTCTGCCCATCTGTAGGTACACAGAGCTCAGCGAATATGTTCT 1815  
 Db 483 ThrGlySerAsnAspLeuProValLeuPheGluLeuGlyGlu---GlySerTyrPheSer 501  
 QY 1816 GTCCCAATCATTTCTTTGGAGATACCATCTCTGTCTCAGCTTACAGTATATCACCACT 1875  
 Db 502 GluGlyAspGlyPheAlaGluAspProThr-----IleSerLeuLeuThrGly 517  
 QY 1876 AGTTCTATGACCATGCTGCCCCCAAGGGCGAGAGCTGCTAGTG 1917  
 Db 518 SerGlu-----ProGlyAlaLysAspProThrVal 528

RESULT 10  
 US-09-513-783A-176  
 ; Sequence 176, Application US/09513783A  
 ; Patent No. 6416959  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Giuliano, Kenneth A.  
 ; APPLICANT: Kapur, Ravi  
 ; TITLE OF INVENTION: A System for Cell Based Screening  
 ; FILE REFERENCE: 97-022-L1  
 ; CURRENT APPLICATION NUMBER: US/09/513,783A  
 ; CURRENT FILING DATE: 2000-02-25  
 ; NUMBER OF SEQ ID NOS: 180  
 ; SOFTWARE: Patent In Ver. 2.0  
 ; SEQ ID NO 176  
 ; LENGTH: 783  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSF1  
 US-09-513-783A-176

Alignment Scores:  
 Pred. No.: 0.00147 Length: 783  
 Score: 140.50 Matches: 111  
 Percent Similarity: 34.24% Conservative: 65  
 Best Local Similarity: 21.60% Mismatches: 185  
 Query Match: 2.39% Indels: 153  
 DB: 4 Gaps: 26

US-10-007-270-1 (1-3330) x US-09-513-783A-176 (1-783)

QY 700 CATTCAAA-----GACTTGGCGAGTATTCTAAGAAAACCCCTCAGAGAGCAAAATCAA 753  
 Db 314 TyrPheLysHisAsnAsnMetalAserPheValArgGlnLeuAsnMetTyrGlyPheArg 333  
 QY 754 GATGTGGCCACAGCTCTCACTTGGGCTTCCCTCTCACTCT-----GATGACACCCCTC 807  
 Db 334 LysValValHisIleGluGlnGlyLeu---ValLysProGluA-rgAspAspThrGlu 352  
 QY 808 CTC-----AATGAAATTCGTGATAATACATCAACGAC 840  
 Db 353 PheGlnHisProCysPheLeuArgGlyGlnGlnLeuLeuGluAsnIleLysArgLys 372  
 QY 841 ACCAAGATGCTCTACACAGAAAGAAACAGAAATTCGTGTGTGGAGGAGCAGAGGGTG 900  
 Db 373 ValThrSerValThrLeuLysSerGluAspIleLysIleArgGlnAspSerValThr 392



```
QY 901 GAGTCAGCGTCTCTCTGTTAAACACAGAGTTCAAGCAGAGAGCTCGTCACTCCACG--- 957
Db 393 LysLeuLeuThrAspValGlnLeuMetLysGlyLysGlnCysMetAspSerLysLeu 412
QY 958 -----TCCCATATTACAGGAGCTAGCAGGAAGTCCCAA--- 993
Db 413 LeuAlaMetLysHisGluAenGluAlaLeuTrpArgGluValAlaSerLeuArgGlnLys 432
QY 994 ---CTTCAGATGCAAAAGATATTAAAGAACTTCCAGATTC-----AAA 1035
Db 433 HisAlaGlnGlnGlnLysValAlaLysLeuIleGlnPheLeuIleSerLeuValGln 452
QY 1036 AAAATCCATGCTTAGGATTTAGACCAAG-----AAAGAAAAGATGCTCA 1083
Db 453 SerAsnArgLysLeuGlyValLysArgLysIleProLeuMetLeuAsnAspSerGlySer 472
QY 1084 AGTCCACAGAGATGCAACTTACGCCATCTTTAAGAGACACAGTGCAGAGCAAAAGC 1143
Db 473 AlaHisSerMetProLysTyrSerArgGlnPheSerLeuGluHisValHisGlySerGly 492
QY 1144 CTGCAAGTGCCTCTGCTCTTTGATTCACAAATAATGAAGTGAAGTCTATCAT 1203
Db 493 Pro-----TyrSerAlaProSerProAlaTyrSerSerSerLeuTyr--- 507
QY 1204 GGAACCATGGAGGACAGCAACCAATCTATCTCACAGCTACAGACCTCAAAAGG 1263
Db 508 -----AlaProAspAlaValAlaSerSerGlyProIle----- 518
QY 1264 CTGATCAGCAAAAGCACTAGAGGAAGACAATCTTGGATGGGGACAATTCAGTTCAC 1323
Db 519 -----IleSerAspIle 522
QY 1324 GATGAATTCGTGATCATCTGCACCCCTTTGGTCT---GACACCCAAATCAGAGTGC 1380
Db 523 ThrGluLeuAlaProAlaSerProMetAlaSerProGlyGlySerIleAspGluArgPro 542
QY 1381 ACATCTTTTGTGTATAACAGAGAGTCTACTTTCAGTCCAGACTCTCTCTCTGGA 1440
Db 543 LeuSerSerSerProLeuValArg-----ValLysGluGluProProSerPro 558
QY 1441 CCCAGCTTGACAGTGGACGGACGAGCATGGTCTACCT---GACACTTCT 1491
Db 559 ProGlnSerProArgValGluGluAlaSerProGlyArgProSerSerValAspThrLeu 578
QY 1492 TGGTCTCCACCTGCTATGCCCTCTACCTCCTGTCAGAGCTCCACTTCTTTATGCA 1551
Db 579 LeuSerProThrAlaLeuIleAspSerIleLeuArgGluSerGluPro-----AlaPro 596
QY 1552 TCAGCATCTTCTCTGACTGAT---CAAGGCACACAGATACAATG----- 1596
Db 597 AlaSerValThrAlaLeuThrAspAlaArgGlyHisThrAspThrGluGlyArgProPro 616
QY 1596 ----- 1596
Db 617 SerProProThrSerThrProGluLysCysLeuSerValAlaCysLeuAspLysAsn 636
QY 1597 -----CCCACTGAC-----CAGCAATGCTA 1617
Db 637 GluLeuSerAspHisLeuAspAlaMetAspSerAsnLeuAspAsnLeuGlnThrMetLeu 656
QY 1617 ----- 1617
Db 657 SerSerHisGlyPheSerValAspThrSerAlaLeuLeuAspLeuPheSerProSerVal 676
QY 1618 ---GTACACGGGTACCATCCACCGAGTGATTATTCTGCAATCAGC---CAACTGGCT 1671
Db 677 ThrValProAspMetSerLeuProAspLeuAspSerSerSerLeuAlaSerIleGlnGluLeu 696
QY 1672 CTGGGATTTTCATCCACCTGCATCTTCAGATGACAGCCGATCAAGTGCA---GGTGGC 1728
Db 697 LeuSerProGlnGluProProArgProProGluAlaGluAsnSerSerProAspSerGly 716
```

```
QY 1729 GAAGATATGGTCAGACAC-----CTAGAT-----GAA 1755
Db 717 LysGlnLeuValHisTyrThrAlaGlnProLeuPheLeuLeuAspProGlySerValAsp 736
QY 1756 ATGGATCTGTGACACACTCTCTGCCCATCTCTGAGGTACAGAGCTCAGCGAATATGTTCT 1815
Db 737 ThrGlySerAsnAspLeuProValLeuPheGluLeuGlyGlu---GlySerTyrPheSer 755
QY 1816 GTCCACGATCATTTCTGCGAGATACCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1875
Db 756 GluGlyAspGlyPheAlaGluAspProThr-----IleSerLeuLeuThrGly 771
QY 1876 AGTTTATGACCATTCGCCCCCAAGGGCCGAGAGCTGGTGTAGTG 1917
Db 772 SerGlu-----ProProLysAlaLysAspProThrVal 782

RESULT 11
US-08-616-844-40
; Sequence 40, Application US/08616844
; Patent No. 5849578
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A.
; TITLE OF INVENTION: COMPOSITION AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/616,844
; FILING DATE: 15-MAR-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,654
; FILING DATE: 09-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-053
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1481 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-616-844-40

Alignment Scores:
Pred. No.: 0.00281 Length: 1481
Score: 139.00 Matches: 180
Percent Similarity: 32.48% Conservative: 113
Best Local Similarity: 19.96% Mismatches: 383
Query Match: 2.37% Indels: 226
```

[illegible]

```

Db 1014 eulleProLeuThrSerValProThrSerAlaLysGluMetThrThrLysLeu----- 1031
QY 2147 CTGTGACACGGCTCTTGGAGATTTCGTCTGCTGAGCCCAACTCCATCTGAAA 2206
Db 1032 -----GlyValThrAlaGluThrProAla----- 1040
QY 2207 TAGACAGCTACTCTCTCAACATTGAACAGCTGATCAA----- 2244
Db 1041 -----SerArgSerLeuGlyThrSerProSerProGlnThrThrValValSerThrAlaG 1059
QY 2244 ----- 2244
Db 1059 luAspLeuAlaProLysSerAlaThrPheAlaValGlnSerSerThrGlnSerProThrT 1079
QY 2245 -----GCAGATCCCTGCAAGTTCCTGCTGCGCGCAATTGCGCC 2284
Db 1079 hrLeuSerSerAlaSerValAsnSerCysAlaValAsnProCysLeuHisAsnGlyG 1099
QY 2285 ANTGTGTAAGAACGACGACTGAGGAAGCGGAGTGCCTGCAACACGAGATATGACA 2344
Db 1099 luCysValAlaAspAsnThrSerArgGlyTyrHisCysArgCysProProSerTrp----- 1117
QY 2345 GCCAGGAGCTGAGGCTGCTGGAACACGAGCTCTGT-----GCCCTGGCA 2392
Db 1118 --GlnGlyAspCysSerValAspValAsnGluCysLeuSerAsnProCysProSerT 1137
QY 2393 CAAGGAATGCGAGTCTCTCAGGGAAGCGAGCTCCATGAGCTGGCA----- 2442
Db 1137 hrAlaThrCysAsnAsnThrGlnGlySerPheIle---CysLysCysProValGlyTyrG 1156
QY 2443 -----GATCCTCTGAAATCAAGATACAAACTAGTGTAAAGATTCCAAATCAAC 2497
Db 1156 InLeuGluLysGlyIleCysAsnLeuValArgThrPheValThrGluPheLysLeuLysA 1176
QY 2498 AAAAT-----ACAGGTAATCATGTAAGAAATCTGAATTACTGACCGCTAGAATATG 2551
Db 1176 rgThrPheLeuAsnThrThrValGluLys---HisSerAspLeuGlnValGluAsnG 1195
QY 2552 AA 2553
Db 1195 lu 1195

```

```

RESULT 12
US-08-599-654-40
; Sequence 40, Application US/08599654
; Patent No. 5882925
; GENERAL INFORMATION:
; APPLICANT: PALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/599,654
; FILING DATE: 09-FEB-1996
; CLASSIFICATION: 800
; PRIORITY INFORMATION:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIORITY INFORMATION:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-041
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9030
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1481 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-599-654-40
Alignment Scores:
Pred. No.: 0.00281 Length: 1481
Score: 139.00 Matches: 180
Percent Similarity: 32.48% Conservative: 113
Best Local Similarity: 19.96% Mismatches: 383
Query Match: 2.37% Indels: 226
DB: 2 Gaps: 39
US-10-007-270-1 (1-3330) x US-08-599-654-40 (1-1481)
QY 242 ATCTCCCAAGAAATGAACAACCTGAA---AGTACTGAAAAAATGTACAAATGTCACT 298
Db 390 AsnProGlyAspGluGluPheIleGluProSerThrGluAsnGluPheGlyLeuThrSer 409
QY 299 ATGACAGCAATATTCAT---TTGGCAAGCATCGAACAAAAAGATCCGCATT 349
Db 410 LeuArgTrpGlnAsnAspSerProThrPheGlyGluHisGlnLeuAlaSerSerSerGlu 429
QY 350 TTCCCAACGGGGCTTAAAGTCTGTCACAGGAATCCATGAAACAGATTTAGACAGCTT 409
Db 430 ValGlnAsnGlySerProMetSerGlnThrGluThrVal-SerArg----- 444
QY 410 CAGCTTATTATAGATGTGTGAGAGAGCATGATGGGAAGCATGATGGATCTTT 469
Db 445 -----SerValAlaProMetArgGlyGlyGluIleThrAlaHi 457
QY 470 CTGGATCGCA-----TCCTGACACAGGGGAATATCAGGCTGGTTCAGCATCTGCCAG 523
Db 457 strPLeuLeuThrAsnSerThrThrSerAlaAspValThrGlySerSerAlaSerTyPr 477
QY 524 CAGGAGACCTCTGCTCTCT---TTGACATTGGAAAAAATTCAGCAATTCACAGG----- 575
Db 477 oGluGlyValAsnAlaSerValLeuThrGlnPheSerAspSerThrValGlnSerGlyG 497
QY 576 -AGCAGCTGATCTCTCCAGCAGAGATAAAACAGAGAGAGTTTCCCTGACAGCAAAAGAT 634
Db 497 ySerHisThrAlaLeuGlyAspArgSerTySerSerSerSerSerSerSerSerSer 516
QY 635 GAAATATCTGACAGAGAGACATTTGGGAGAGCTGCTGTAACCATTTGTCATTTCAACAGCA 694
Db 517 -----GluSerLeuAsnSerSerAlaProArgGlyGluAr 528
QY 695 ATCTACATTCAAAGACTTGGGAGTATTCTTAAGAAACCTCAGAGAGCAAAATTCAG 754
Db 528 gSerThrLeuGlu-Asp-----SerArgGluProGlyGlnAlaLeuGlyAspS 544
QY 755 ATGTGCAACGCTC-----TCACTTGGGCGCTTTCCCTCTCA 790
Db 544 exSerAlaAsnAlaGluAspArgThrSerGlyValProSerLeuGlyThrHisThrLeuA 564
QY 791 CTCCT-----GATGACACCTCTCTCAATGAATTTCTCGATAATACACTCAACG 838
Db 564 laThrValThrGlyAsnGlyGluArgThrLeuArgSerVal-----ThrLeuThrA 581
QY 839 ACACCAAGATGCTTACACAGAA-----AGAG 865

```

581 snThrSerMetSerThrThrThrSerGlyGluAlaGlySerProAlaAlaAlaMetProGlnG 601  
QY 866 AAACAGAAATTGCGTCTGCTGGAGGACGAGGGTGGAGCTCAGCGCTCTCTGGTAACC 925  
Db 601 luthrGlyAlaSerLeuHis----- 608  
QY 926 AGAAGTTCAAGCGCAGACTCGCTGACTCCCGTACCTCCAGTCCCCCATATTACCAGGAGCTAGCAGAA 985  
Db 609 -----ValAsnValThrAspMetGlyLeuValSerArgSerLeuAlaAlaAs 625  
QY 986 AGTCCCACTTCAGATCCAAAGATATTTAAGAACTTCAGGATTCAAAANAATCCATG 1045  
Db 625 erSerAlaLeuGlyAlaGlyLe-----SerTyrGlyGlnValArg 640  
QY 1046 TGTAGGATTTAGACCAAGAAGAAAGAAAGATCGCTCAAGCTCCACAGAGATCGACATTA 1105  
Db 640 lyThrAlaIleGluAlaGthrSerSerAspHisThrAspHisThr-----TyrLeuS 658  
QY 1106 CGGCATCTTTAAGACAGAC-----AGTGCAGAAAGCAAAAGCCCTCGAA 1150  
Db 658 erSerThrPheThrLysGlyGluArgAlaLeuLeuSerIleThrAspAsnSerSerSerS 678  
QY 1151 GTCACTCTCTCTTTTGATTCACAAAATTCAAAGTCAGGAAGTCATCATGAGCAACA 1210  
Db 678 erAspIleValGluSerSerSerTyrIleLysIleSerAsnSerSerHisSerGluT 698  
QY 1211 TGGAGGAGCAAGCAACCAAGAAATCTATCTCACAGCTACAGACTCCAAAGCGCTGATCA 1270  
Db 698 yrSerSerPheSerHisAlaGln-----ThrGluArgSerAsnIleSerS 713  
QY 1271 GCAAGCACTAGAGGAGAAACAATCTTTGGATGTGGGCAATTCAGTTCACTGATGAAA 1330  
Db 713 erTyrAspGlyGluTyrAlaGlnProSerThrGluSerProValLeuHisThrSerAsnL 733  
QY 1331 TTGCTGGATCACTGCCAGCCTTTGGT---CCTGCACACCAATCAGAGCTGCCACATCTT 1387  
Db 733 euProSerTyrThrProThrIleAsnMetProAsnThrSerValValLeuAspThrAspA 753  
QY 1388 TTGCTGTTATACAGAGGATCTACTTTGAGT----- 1419  
Db 753 laGluPheValSerAspSerSerSerSerSerSerSerSerSerSerSerS 773  
QY 1420 --CCAGAACTTCTCTGTTGACCCCACTTGAGAGCTGGAGCGAGCAGAGCAT---- 1473  
Db 773 lyProProLeuPro-----LeuProSerValSerGlnSerHisIleuP 788  
QY 1474 -----GGTCTACCTGCACACTTCTTGCTCTCCACCTGTCTATGGCTCTACCTCCCTG 1525  
Db 788 heSerIleLeuProSerThrArgAlaSerValHisLeuLeuLysSerThrSerAspA 808  
QY 1526 CAGAGCTCCACCTTTCTTTATGGCATCAAGCATCTTCTCTGACTGATCAAGGACCA 1585  
Db 808 laSerThrPro-----TyrSerSerProSerProLeuProValSerLeuThrT 825  
QY 1586 CAGATACATGGCCACTCACCACCAATCTAGTACAGGGCTTCACCATCCCAACCAAGTG 1645  
Db 825 hrSerThrSerAla-----ProLeuSerValSerGlnThrThrLeuProGlnSerS 842  
QY 1646 ATTATTCTCCAAATCAGCAACTGGCTCTGGGAATTTCCATCCACTGCATCTTCAGATG 1705  
Db 842 erSerThrProValLeuProArgAla-----ArgGluThrProValThrSerPheG 859  
QY 1706 ACAGCGGATCAAGTGCAGGTGGCAGAGATATGGTCACACACCTAGATGAATGATCTGT 1765  
Db 859 lnThrSerThrMetThrSerPheMetThrMetLeuHisSerSerGlnThrAlaAspLeuL 879  
QY 1766 CTGACACTCTGCCCATCTGAGGTACGAGAGCTCAGCGAA-----TATG 1810  
Db 879 ysserGlnSerThrProHisGlnGluLysValIleThrGluSerLysSerProSerLeuV 899  
QY 1811 TTCTCTCCCGCATCTTTCTGGAGGATACCACTCTCTCTCAGCTTTTCAGTATATCA 1870  
Db 899 alserLeuProThrGluSerThrLysAlaValThrThrAsnSerProLeuProProSerL 919

[illegible]

RESULT 13

```

US-08-944-868A-40
; Sequence 40, Application US/08944868A
; Patent No. 6018025
; GENERAL INFORMATION:
;   APPLICANT: FALB, DEAN A
;   TITLE OF INVENTION: COMPOSITIONS AND METHODS OF USE
;   TITLE OF INVENTION: TREATMENT AND PREVENTION OF

```



QY 1526 CAGAAGCTCCACTTCTTTATGGCATCAGCATCTCTCTGACTGATCAAGGCACCA 1585  
 Db 808 laSerThrPro-----TrrSerSerProSerProLeuProValSerLeuThrt 825  
 QY 1586 CAGATACAAATGCCACTGACACAGATGCTAGTACAGGGCTCACCATCCCAACGATG 1645  
 Db 825 hrSerThrSerAla-----ProLeuSerValSerGlnThrLeuProGlnSerS 842  
 QY 1646 ATTATTCGCAATCAGCAACTGGCTCTGGGAATTTTCATCCCTGCTGATGATG 1705  
 Db 842 erSerThrProValLeuProArgala-----ArgGluThrProValThrSerPheG 859  
 QY 1706 ACAGCCCATCAAGTGCAGGTGGCGAGATATGTCAGACACTAGATGAATGATCTGT 1765  
 Db 859 lnThrSerThrMetThrSerPheMetThrMetLeuHisSerSerGlnThrAlaAspLeuL 879  
 QY 1766 CTGACACTCTCGCCCATCTGAGGTACAGACTCAGCGAA-----TATG 1810  
 Db 879 ysSerGlnSerThrProHisGlnGluLysValIleThrGluSerLysSerProSerLeuV 899  
 QY 1811 TTTCTGTCCTCAGATCATTTCTTGAGGATACACTCTGCTCAGCTTTTACAGTATATCA 1870  
 Db 899 aSerLeuProThrGluSerThrLysAlaValThrThrAsnSerProLeuProProSerL 919  
 QY 1871 COACTAGTTCTATGACCAATTCGCCCAAGCGCGAGAGCTGCTAGTGTCTTCAGTCTGC 1930  
 Db 919 euThrGluSerSerThr-----GluGlnThrLeuProAlaThrSerThrA 934  
 QY 1931 GTGTGTCTACATGGCC-----TTCCTCAACGACTGTTTCAACAGAGC----- 1974  
 Db 934 enLeuAlaGlnMetSerProThrPheThrThrThrIleLeuLysThrSerGlnProLeuM 954  
 QY 1975 -----TCTCTGGAGTACGAGCTCTGGAGCAAC 2002  
 Db 954 etThrThrProGlyThrLeuSerSerThrAlaSerLeuValThrGlyProIleAlaValG 974  
 QY 2003 AATTCACA-----CAGCTGCTGTGTTCCATCTA- 2031  
 Db 974 lnThrAlaGlyLysGlnLeuSerLeuThrHisProGluIleLeuValProGlnIleS 994  
 QY 2032 -----CGATCCCAATTTACAGGATTTAAGCACTGTAATCTTCAAGAAACGGA 2086  
 Db 994 erThrGluGlyLysThrGluArgAsnArgValIleValAspAlaThrThrGlyL 1014  
 QY 2087 GTGTGATTTGTAATGCAAAATGAGTTTGTAGTCTGCGGTATTAACCTCACAAGG 2146  
 Db 1014 eulleProLeuThrSerValProThrSerAlaLysGluMetThrThrLysLeu----- 1031  
 QY 2147 CTGTGCACGGGCTCTGGAGGATTTTCTGCTGCTGCGAGCCCAACACTCCATCTCGAA 2206  
 Db 1032 -----GlyValThrAlaGluThrSerProAla----- 1040  
 QY 2207 TAGACAGCTACTCTCTCACTGATGACAGCTGATCAA----- 2244  
 Db 1041 -----SerArgSerLeuGlyThrSerProSerProGlnThrThrValValSerThrAlaG 1059  
 QY 2244 ----- 2244  
 Db 1059 luAspLeuAlaProLysSerAlaThrPheAlaValGlnSerSerThrGlnSerProThrT 1079  
 QY 2245 -----GCAGATCCCTGCAAGTTCTGCTGCGCTGGCGGGAATTTGCC 2284  
 Db 1079 hrLeuSerSerSerAlaSerValAsnSerCysAlaValAsnProCysLeuHisGlyG 1099  
 QY 2285 AATGTGTAAGNACGACGAGCTGAGGAGCGAGTGTGCTGCAAAACAGGATATGACA 2344  
 Db 1099 luCysValAlaAspAsnThrSerArgGlyTyriHisCysArgProProSerTrp----- 1117  
 QY 2345 GCCAGGGAGGCTGACGCTCTGGAACACGAGCTCTGT-----GGCCTTGCCA 2392  
 Db 1118 --GlnGlyAspAspCysSerValAspValAsnGluCysLeuSerAsnProCysProSerT 1137  
 QY 2393 CAAGGAATGCGAGTCTCTCCAGGAAAGGAGCTCCATGCGAGTGTCCCA----- 2442

Db 1137 hrAlaThrCysAsnAsnThrGlnGlySerPheIle-----CysLysCysProValGlyTyrg 1156  
 QY 2443 -----GATCACTCTGAAATCAAGCATACAACTAGTGTAAAGTTCCAAATCAAC 2497  
 Db 1156 lnLeuLysGlyIleCysAsnLeuValArgThrPheValThrGluPheLysLeuLysA 1176  
 QY 2498 ABAAT-----AACAGGTAATCAGTAAGAAATTTCTGAATTCAGCCGTAGATATG 2551  
 Db 1176 rgThrPheLeuAsnThrThrValGluLys--HisSerAspLeuGlnGluValGluAsnG 1195  
 QY 2552 AA 2553  
 Db 1195 lu 1195  
 RESULT 14  
 US-08-944-423A-40  
 ; Sequence 40, Application US/08944423A  
 ; Patent No. 6020463  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FALB, DEAN A  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
 ; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE  
 ; NUMBER OF SEQUENCES: 54  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSER: PENNIE & EDMONDS  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: WINDOWS 95  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/944,423A  
 ; FILING DATE: 06-OCT-1997  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/599,654  
 ; FILING DATE: 09-FEB-1996  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/485,573  
 ; FILING DATE: JUN-07-1995  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/386,844  
 ; FILING DATE: 10-FEB-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: CORUZZI, LAURA A  
 ; REGISTRATION NUMBER: 30,742  
 ; REFERENCE/DOCKET NUMBER: 7853-105  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 790-9090  
 ; TELEFAX: (212) 869-8864  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 40:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1481 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS:  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; US-08-944-423A-40  
 Alignment Scores:  
 Pred. No.: 0.00281  
 Score: 139.00  
 Percent Similarity: 32.48%  
 Best Local Similarity: 19.96%  
 Query Match: 2.37%  
 Gaps: 3  
 Length: 1481  
 Matches: 180  
 Conservative: 113  
 Mismatches: 383  
 Indels: 226  
 Gaps: 39



|  |    |      |   |      |
|--|----|------|---|------|
|  | QY | 2147 | CTGTGCACGGGTCCTTGAGGAATTTTGGTTCTGTCTCGCAGCCCAACAACCTCATCTCGAA                                     | 2206 |
|  | Dd | 1032 | :   <br>-----GlyAlaThrAlaGlutyrSerProA-   | 1040 |
|  | QY | 2207 | TAGACAGCTACTCTCTCAAACATTGAACCAGCTGATCAA--   | 2244 |
|  | Dd | 1041 | :   <br>----SerArgSerLeuGlyThrSerProSerProGlnThrThrValValSerThrAlag                               | 1059 |
|  | QY | 2244 | -- -- -- --   | 2244 |
|  | Dd | 1059 | luAspLeuAlaProLysSerAlaThrPheAlaValGlnSerSerThrGlnSerProThr                                       | 1079 |
|  | QY | 2245 | -----GCCAGATCCCTCGCAAGTCTCTGGCTCGGGCGAATTGGCC   | 2284 |
|  | Dd | 1079 | :    :   <br>hrLeuSerSerSerAlaSerValAsnSerCyAlaValAsnProCysLeuHisAsnGlyg                          | 1099 |
|  | QY | 2285 | AATGTGTAAGAACGACGGACTTAGAAGCGAGTCTCGCTGCAAAACGAGTAGTACA   | 2344 |
|  | Dd | 1099 | ::::  :::<br>luCyValAlaAspAnthrSerzArgIyTyHicCysArzCysProSerTrp----                               | 1117 |
|  | QY | 2345 | GCCAGGGAGCCTGTGGACCGAGCTCTGMAACCGAGGCTCTGT-----GGCCCTGGCCA  | 2392 |
|  | Dd | 1118 | - - - GlnGIyAspApCysserValAspValasngInuCysLeuSerzAsnPcCsProSert                                   | 1137 |
|  | QY | 2393 | CMAAGAATCGAGGTCCTCCAGCGAAAAGGAGCTCCATCGAGGTTCGCA- ----  | 2442 |
|  | Dd | 1137 | <br>hrAlaThrCysAenAsnthrgLnGLySerPhelIE---CyslScypProValGilytyrG                                  | 1156 |
|  | QY | 2443 | -----GATCCTCTGMAATOCAGCATACAAAACTAGTGTATAAAGTTC AAAATCAAC   | 2497 |
|  | Dd | 1156 | :    :   <br>lmLeuGluLySLyglileCysasnLeuValargThrPheValThrGluPheLysLeulySA                        | 1176 |
|  | QY | 2498 | AAAA-----AACRAGGTATCAGTAAAGAAATCTCGAATTACTCACCGCTAGAAATG  | 2551 |
|  | Dd | 1176 | :    :    :    :    :    :    :   <br>rgThrPheLeuAsnthrThrValGluLys---HisSeraspLeuglInuValGluAsng | 1195 |
|  | QY | 2552 | AA 2553   |      |
|  | Dd | 1195 | lu 1195   |      |

## RESULT 15

US-08-944-496-40  
; Sequence 40, Application US/08944496  
; Patent No. 6124433  
; GENERAL INFORMATION:  
; APPLICANT: FALB, DEAN A  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TITILE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/944,496  
; FILING DATE: 06-OCT-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/599,654  
; FILING DATE: 09-FEB-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,573  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:



```

Db      899  alSerLeuProThrGluSerThrLysAlaValThrThrAsnSerProLeuProProSerL 919
      |||||:|||||
      :|||
      |||
QY 1871  COACTAGTTCTATGACCAATGCCCCCAAGGCGGAGAGCTGGTAGTGTCTTCAGTCTGC 1930
      |||||
      |||||
Db  919  euThrGluSerThr-----GluGlnThrLeuProAlaThrSerThrA 934
      :|||
      :|||
      :|||
QY 1931  GTGTGTCTTAACATGCCC-----TTTCCCAACGACCTGTTCAACAAGAC----- 1974
      :|||
      :|||
      :|||
Db  934  snLeuAlaGlnMetSerProThrPheThrThrThrIleLeuLysThrSerGlnProLeuM 954
      :|||
      :|||
      :|||
QY 1975  -----TCTCTGGAGTACCAGACTCTGGAGCAAC 2002
      :|||
      :|||
      :|||
Db  954  etThrThrProGlyThrLeuSerSerThrAlaSerLeuValThrGlyProIleAlaValG 974
      :|||
      :|||
      :|||
QY 2003  AATTCACA-----CAGCTGTGGTGGTTCATATCTTA----- 2031
      :|||
      :|||
      :|||
Db  974  lnThrThrAlaGlyLysGlnLeuSerLeuThrHisProGluIleLeuValProGlnIle 994
      :|||
      :|||
      :|||
QY 2032  -----CGATCCAACTTACAGGATTATAGCAACTTGAATATCTTAACCTTCAGAAACGGGA 2086
      :|||
      :|||
      :|||
Db  994  erThrGluGlyLysThrThrGluArgAsnArgValIleValAspAlaThrThrGlyL 1014
      :|||
      :|||
      :|||
QY 2087  GTGTGATGTGAATAGCAAAATCAAGTTTGCTAGTCTGTGCGGTATTAACCTTCACCAAG 2146
      :|||
      :|||
      :|||
Db  1014  euIleProLeuThrSerValProThrSerAlaLysGluMetThrThrLysLeu----- 1031
      :|||
      :|||
      :|||
QY 2147  CTGTGCACGGGCTCTGGAGGATTTCTGCTGTCGACGCCCAACAACATCCATCTCGAAA 2206
      :|||
      :|||
      :|||
Db  1032  -----GlyValThrAlaGluTyrSerProAla----- 1040
      :|||
      :|||
      :|||
QY 2207  TAGACAGCTACTCTCTCAACATTGAACCACTGATCAA----- 2244
      :|||
      :|||
      :|||
Db  1041  -----SerArgSerLeuGlyThrSerProSerProGlnThrThrValValSerThrAlaG 1059
      :|||
      :|||
      :|||
QY 2244  ----- 2244

Db  1059  luAspLeuAlaProLysSerAlaThrPheAlaValGlnSerSerThrGlnSerProThr 1079
      :|||
      :|||
      :|||
QY 2245  -----GCAGATCCCTGCAGTTCCTGGCTCGCGCGAATTGCCC 2284
      :|||
      :|||
      :|||
Db  1079  hrLeuSerSerSerAlaSerValAsnSerCysAlaValAsnProCysLeuHisAsnGlyG 1099
      :|||
      :|||
      :|||
QY 2285  AATGTGTAAAGAACGAACGACTGAGGACGGAGTGTGCTGCACCAACCGAGATATGACA 2344
      :|||
      :|||
      :|||
Db  1099  luCysValAlaAspAsnThrSerArgGlyTyrHisCysArgCysProProSerTrp----- 1117
      :|||
      :|||
      :|||
QY 2345  GCCAGGGGACCTTGGACGGTCTGGAACAGGCTCTGT-----GGCCCTGGCA 2392
      :|||
      :|||
      :|||
Db  1118  --GlnGlyAspAepCysSerValAspValAsnGluCysLeuSerAsnProCysProSer 1137
      :|||
      :|||
      :|||
QY 2393  CAAGAAGAAATCGAGGTCTCCAGGGAAGGAGCTCCATGAGGTGCA----- 2442
      :|||
      :|||
      :|||
Db  1137  hrAlaThrCysAsnAsnThrGlnGlySerPheIle--CysLysCysProValGlyTyrG 1156
      :|||
      :|||
      :|||
QY 2443  -----GATCACTCTGAAAATCAAGCATACAAAACCTAGTGTATAAAGTTGCCAAATCAAC 2497
      :|||
      :|||
      :|||
Db  1156  lnLeuGlnLysGlyIleCysAsnLeuValArgThrPheValThrGluPheLysLeuLysA 1176
      :|||
      :|||
      :|||
QY 2498  AAAAT-----NACAGGTATCAGTAAAGAAATCTGAAATTACTGACCGTAGAATATG 2551
      :|||
      :|||
      :|||
Db  1176  rgThrPheLeuAsnThrThrValGluLys--HisAspLeuGlnGluValGluAsnG 1195
      :|||
      :|||
      :|||
QY 2552  AA 2553
      :|||
      :|||
      :|||
Db  1195  lu 1195

```

Search completed: March 4, 2004, 18:56:09  
Job time : 109 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 4, 2004, 18:45:37 ; Search time 112 Seconds  
(without alignments)  
12556.078 Million cell updates/sec

Title: US-10-007-270-1

Perfect score: 5875

Sequence: 1 taaaccaagaaggttatctct.....tactatatgacataatcaat 3330

Scoring table: BLOSUM62

|                           |
|---------------------------|
| Xgapop 10.0 , Xgapext 0.5 |
| Ygapop 10.0 , Ygapext 0.5 |
| Fgapop 6.0 , Fgapext 7.0  |
| Delop 6.0 , Delext 7.0    |

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 1619484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DEV=xlp  
-Q/cgn2\_1/USPTO\_spool\_p/US10007270/runat\_04032004\_160718\_5125/app\_query.fasta\_1.3527  
-DB=Published Applications AA -QFMT=fastan -SUFFIX=xapb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pcr -THR\_MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10007270 @CGN 1 1 22 @runat\_04032004\_160718\_5125  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:\*

|  |
|--|
| 1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*   |
| 2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*    |
| 3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*   |
| 4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*   |
| 5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*   |
| 6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*  |
| 7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*   |
| 8: /cgn2_6/prodata/2/pubpaa/US09_PUBCOMB.pep.*   |
| 9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*  |
| 10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.* |
| 11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.* |
| 12: /cgn2_6/prodata/2/pubpaa/US09D_PUBCOMB.pep.* |
| 13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.* |
| 14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.* |
| 15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.* |
| 16: /cgn2_6/prodata/2/pubpaa/US10D_PUBCOMB.pep.* |
| 17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*  |
| 18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*  |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| -----      |       |             |        |    |             |

|    |        |      |      |    |                     |                    |
|----|--------|------|------|----|---------------------|--------------------|
| 1  | 4073   | 69.3 | 797  | 13 | US-10-007-270-2     | Sequence 2, Appli  |
| 2  | 3862   | 65.7 | 771  | 13 | US-10-007-270-28    | Sequence 28, Appli |
| 3  | 3611   | 71.9 | 13   |    | US-10-007-270-4     | Sequence 4, Appli  |
| 4  | 2504.5 | 42.6 | 798  | 13 | US-10-007-270-9     | Sequence 9, Appli  |
| 5  | 1298.5 | 22.1 | 466  | 13 | US-10-007-270-11    | Sequence 11, Appli |
| 6  | 1038   | 17.7 | 198  | 13 | US-10-007-270-6     | Sequence 6, Appli  |
| 7  | 917    | 15.6 | 1241 | 13 | US-10-007-270-17    | Sequence 17, Appli |
| 8  | 841    | 14.3 | 185  | 13 | US-10-007-270-15    | Sequence 15, Appli |
| 9  | 820.5  | 14.0 | 1069 | 13 | US-10-007-270-24    | Sequence 24, Appli |
| 10 | 420.5  | 7.2  | 432  | 13 | US-10-007-270-19    | Sequence 19, Appli |
| 11 | 189.5  | 3.2  | 1235 | 10 | US-09-996-069-10    | Sequence 10, Appli |
| 12 | 189.5  | 3.2  | 1255 | 14 | US-10-171-311-158   | Sequence 158, App  |
| 13 | 189.5  | 3.2  | 1255 | 14 | US-10-171-311-158   | Sequence 158, App  |
| 14 | 188.5  | 3.2  | 515  | 14 | US-10-097-340-212   | Sequence 311, App  |
| 15 | 188.5  | 3.2  | 515  | 14 | US-10-097-340-212   | Sequence 311, App  |
| 16 | 185.5  | 3.2  | 508  | 14 | US-10-057-136-20    | Sequence 156, App  |
| 17 | 185.5  | 3.1  | 475  | 15 | US-10-417-312-1     | Sequence 20, Appli |
| 18 | 169    | 2.9  | 67   | 13 | US-10-007-270-13    | Sequence 1, Appli  |
| 19 | 161.5  | 2.7  | 1905 | 15 | US-10-259-194A-86   | Sequence 86, Appli |
| 20 | 160.5  | 2.7  | 3507 | 15 | US-10-369-493-5784  | Sequence 5784, Ap  |
| 21 | 151.5  | 2.6  | 1161 | 14 | US-10-017-161-2398  | Sequence 2398, Ap  |
| 22 | 151.5  | 2.6  | 1161 | 15 | US-10-292-798-2040  | Sequence 2040, Ap  |
| 23 | 150.5  | 2.6  | 1066 | 14 | US-09-881-752A-166  | Sequence 166, App  |
| 24 | 150    | 2.6  | 1163 | 9  | US-10-029-386-32403 | Sequence 32403, A  |
| 25 | 150    | 2.6  | 1163 | 9  | US-09-893-348-18    | Sequence 18, Appli |
| 26 | 148.5  | 2.5  | 947  | 14 | US-10-293-822-1     | Sequence 1, Appli  |
| 27 | 148    | 2.5  | 1490 | 15 | US-10-177-293-86    | Sequence 86, Appli |
| 28 | 148    | 2.5  | 1367 | 9  | US-10-435-696-31    | Sequence 31, Appli |
| 29 | 147    | 2.5  | 1367 | 9  | US-09-801-368-108   | Sequence 108, App  |
| 30 | 146.5  | 2.5  | 2409 | 14 | US-10-177-293-90    | Sequence 90, Appli |
| 31 | 145    | 2.5  | 630  | 14 | US-10-207-655-69    | Sequence 69, Appli |
| 32 | 145    | 2.5  | 1957 | 15 | US-10-369-493-2070  | Sequence 2070, Ap  |
| 33 | 144.5  | 2.5  | 1431 | 14 | US-10-102-555-2     | Sequence 2, Appli  |
| 34 | 144    | 2.5  | 1881 | 14 | US-10-032-585-7646  | Sequence 7646, Ap  |
| 35 | 143    | 2.4  | 321  | 9  | US-09-925-301-861   | Sequence 861, App  |
| 36 | 141.5  | 2.4  | 2344 | 9  | US-09-815-242-12713 | Sequence 12713, A  |
| 37 | 141    | 2.4  | 3562 | 15 | US-10-341-434-109   | Sequence 109, App  |
| 38 | 140.5  | 2.4  | 529  | 14 | US-10-046-420-2     | Sequence 2, Appli  |
| 39 | 140.5  | 2.4  | 783  | 14 | US-10-100-957A-176  | Sequence 176, App  |
| 40 | 140.5  | 2.4  | 1259 | 14 | US-10-032-585-7120  | Sequence 7120, Ap  |
| 41 | 139.5  | 2.4  | 726  | 15 | US-10-012-697-1480  | Sequence 1480, Ap  |
| 42 | 139.5  | 2.4  | 1928 | 15 | US-10-369-493-22025 | Sequence 22025, A  |
| 43 | 139    | 2.4  | 1052 | 15 | US-10-094-749-2050  | Sequence 2050, Ap  |
| 44 | 139    | 2.4  | 1481 | 9  | US-09-371-900-40    | Sequence 40, Appli |
| 45 | 139    | 2.4  | 1481 | 9  | US-09-924-417-60    | Sequence 60, Appli |

## ALIGNMENTS

RESULT 1  
US-10-007-270-2  
Sequence 2, Application US/10007270  
Publication No. US20020160954A1  
GENERAL INFORMATION:  
APPLICANT: Hageman, Gregory S.  
APPLICANT: Kuehn, Markus H.  
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS FOR OCULAR ABNORMALITIES  
FILE REFERENCE: 020618-000120US  
CURRENT APPLICATION NUMBER: US/10/007,270  
PRIOR FILING DATE: 2001-11-08  
PRIOR APPLICATION NUMBER: US 09/430,195  
PRIOR FILING DATE: 1999-10-29  
PRIOR APPLICATION NUMBER: US 09/183,972  
PRIOR FILING DATE: 1998-10-29  
NUMBER OF SEQ ID NOS: 37  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 797  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Human IPM 150 amino acid sequence, isoform A

## US-10-007-270-2

## Alignment Scores:

Pred. No.: 0  
 Score: 4073.00  
 Percent Similarity: 97.31%  
 Best Local Similarity: 97.19%  
 Query Match: 69.33%  
 DB: 13  
 Length: 797  
 Matches: 796  
 Conservative: 1  
 Mismatches: 0  
 Indels: 22  
 Gaps: 1

US-10-007-270-1 (1-3330) x US-10-007-270-2 (1-797)

QY 131 ATGATTTTGGAACTAGAGAGCTATTTTCTTTTGGATTTTCTCCAGTCAAGGA 190  
 Db 1 MetTyrLeuGluThrArgArgAlaIlePheValPheThrIlePheLeuGlnValGlnGly 20  
 QY 191 ACCAAGATATCTCCATTAACTATACATCTCAATCTGAACTAAAGACATAGACATCCCCCA 250  
 Db 21 ThrLysAspIleSerIleAsnIleTyrHisSerGluThrLysAspIleAspAsnProPro 40  
 QY 251 AGAAATGAAACAACTGAAAGTACTGAAATAATGTAACAAATGTCAACTATGAGCAATA 310  
 Db 41 ArgAsnGluThrThrGluSerThrGluLysMetTyrLysMetSerThrMetArgArgIle 60  
 QY 311 TTCGATTTGGCAAGCATCGAACAAAAGATCGCATTTTCCCAAGCGGGTAAAGTC 370  
 Db 61 PheAspLeuAlaLysHisArgThrLysArgSerAlaPhePheProThrGlyValLysVal 80  
 QY 371 TGCTCCACAGGAATCCATGAACAGATTTTAGACAGTCTTCAAGCTTATTATAGATTGAGA 430  
 Db 81 CysProGlnGluSerMetLysGlnIleLeuAspSerLeuGlnAlaTyrTrgArgLeuArg 100  
 QY 431 GTGTGTCAGAACGACTATGGAGAGCATATCCGATCTTTCTGAGATCGCATCCCTGACACA 490  
 Db 101 ValCysGlnGlnAlaValTrpGluAlaTyrArgIlePheLeuAspArgIleProAspThr 120  
 QY 491 GGGGAATATCAGAGCTGGGTGAGCATCTCCAGCAGAGAGCTTTCTGAGATCGCATCCCTGACACA 490  
 Db 121 GlyGluTyrGlnAspTrpValSerIleCysGlnGlnGluThrPheCysLeuPheAspIle 140  
 QY 551 GGAATAACTTCAGCAATTCAGAGAGCACTGGATCTTCTCCAGCAGAGATAAACAAG 610  
 Db 141 GlyLysAsnPheSerAsnSerGlnGluHisLeuAspLeuLeuGlnGlnArgIleLysGln 160  
 QY 611 AGAAGTTCTCCCTGACAGAAAAGATGAATATCTCAGAGAGACATTTGGAGAGCTGGT 670  
 Db 161 ArgSerPheProAspArgLysAspGluIleSerAlaGluLysThrLeuGluProGly 180  
 QY 671 GAAACCATTTGTCATTTCAACAGCAATCTACATTTCAAAGACTTGGGCAGTATTCTAAGA 730  
 Db 181 GluThrIleValIleSer--Thr----- 187  
 QY 731 AACCTCAGAGAGCAAAATTCAGAGATGTTGGACGTCATCTGGGCTTTCCCTCTCA 790  
 Db 188 -----AspValAlaAsnValSerLeuGlyProPheProLeu 200  
 QY 791 CTCTGATGACACCTCTCTCAATGAATTTCTCGATAATACACTCAACGACACCAAGATGC 850  
 Db 200 hrProAspAspThrLeuLeuAsnGluIleLeuAspAsnThrLeuAsnAspThrLysMetP 220  
 QY 851 CTACACAGAAAGAAACAAATTCGCTGTGTGGAGGAGCAGAGGCTGGAGCTCAGCG 910  
 Db 220 roThrGluArgGluThrGluPheAlaValLeuGluGluLysValGluLeuSerV 240  
 QY 911 TCTCTCTGTAACCCAGAGTTCAAGGACAGAGCTCGCTGATCTCCAGTCCCAATATTACC 970  
 Db 240 alSerLeuValAsnGlnLysPheLysAlaGluLeuAlaAspSerGlnSerProTyrTrg 260  
 QY 971 AGAGCTAGCAGAAAGTCCCACTTCAGATGCAAAAGATATTAAAGACTTCCAGAT 1030  
 Db 260 InGluLeuAlaGlyLysSerGlnLeuGlnMetGlnLysIlePheLysLysLeuProGlyP 280  
 QY 1031 TCAAAAATCCATGTGTAGATTATAGACAAAGAAAGAAAGATGGCTCAAGCTCA 1090

Db 280 helysLysIleHisValLeuGlyPheArgProLysLysGluLysAspGlySerSerSert 300  
 QY 1091 CAGAGATGCAAACTTACGCCCATCTTTAAGAGACACATGTCAGAGCAAAAGCCCTCCAA 1150  
 Db 300 hrGluMetGlnLeuThrAlaIlePheLysArgHisSerAlaGluAlaLysSerProAlas 320  
 QY 1151 GTGACCTCTCTCTTTTGTATCCAAACAAATTTGAAGTGAAGGAGTATCATCGAACCA 1210  
 Db 320 erAspLeuLeuSerPheAspSerAsnLysIleGluSerGluGluValLysHisGlyThrW 340  
 QY 1211 TGGAGAGACACAGCAACCAACCAATATCTATCTCAGCTACAGACTCAAAAGGTGATCA 1270  
 Db 340 etGluGluAspLysGlnProGluIleTyrLeuThrAlaThrAspLeuLysArgLeuIleS 360  
 QY 1271 GCAAGCACTAGAGAGAAACAATCTTTGGATGTGGGCAATTCAGTTCCTACTGATGAAA 1330  
 Db 360 erLysAlaLeuGluGluGlnSerLeuAspValGlyThrIleGlnPheThrAspGluI 380  
 QY 1331 TTGCTCGATCATCTCCAGGCTTTGTCTCTGACACCAATCAGAGCTGCCACATCTTTTG 1390  
 Db 380 leaGlySerLeuProAlaPheGlyProAspThrGlnSerGluLeuProThrSerPheA 400  
 QY 1391 CTCTTATAACAGAGATGTCTTCTTGTCTCCAGACTTCTCTCTGTTGAAACCCAGCTTG 1450  
 Db 400 laValIleThrGluAspAlaThrLeuSerProGluLeuProProValGluProGlnLeuG 420  
 QY 1451 AGACGTGGACGGAGCAGAGCATGTCTTACTGACACTTCTTGTCTCCACCTGTATGG 1510  
 Db 420 luThrValAspGlyAlaGluHisGlyLeuProAspThrSerTrpSerProProAlaMetA 440  
 QY 1511 CTCTTACCTCTCTGTCCAGAGCTCCACTTCTTTATGGCATCAAGCATCTTCTCTCTGA 1570  
 Db 440 laseThrSerLeuSerGluAlaProPhePheMetAlaSerSerIlePheSerLeuT 460  
 QY 1571 CTGATCAAGGCACACACAGATCAATGCGCACTGACACACAGCAAGCTAGTACAGGGTCA 1630  
 Db 460 hrAspGlnGlyThrThrAspThrMetAlaThrAspGlnThrMetLeuValProGlyLeuT 480  
 QY 1631 CATCTCCACAGCATGATTTCTCCATCAGCCAACTGGCTCTGGGAATTTTCAATCCAC 1690  
 Db 480 hrIleProThrSerAspTyrSerAlaIleSerGlnLeuAlaLeuGlyIleSerHisProp 500  
 QY 1691 CTGCATCTTCAGATGACAGCGCATCAAGTGCAGGTGGCGAAGATATGTCACACACCTAG 1750  
 Db 500 roAlaSerSerAspAspSerArgSerSerAlaGlyGlyGluAspMetValArgHisLeuA 520  
 QY 1751 ATGAAATGGATCTGTGACACTCTCTGCCCATCTGAGGTACAGAGCTCAGCGCAATATG 1810  
 Db 520 epGluMetAspLeuSerAspThrProAlaProSerGluValProGluLeuSerGluTyrV 540  
 QY 1811 TTTCGTCTCCAGATCAATTTCTGGAGATACCACTCTCTGTCTCAGCTTTACAGTATATCA 1870  
 Db 540 alSerValProAspHisPheLeuGluAspThrThrProValSerAlaLeuGlnTyrIleT 560  
 QY 1871 CCCTAGTTCTATACCATTTGCCCCAGGCGCAGAGCTGGTAGTGTCTTCTGCTCTCC 1930  
 Db 560 hrThrSerSerMetThrIleAlaProLysGlyArgGluLeuValValPhePheSerLeuA 580  
 QY 1931 GTGTCTCTAACATGCGCTTCTCCACAGCACTGTTTCAACAGAGCTCTCTGAGTACCGAG 1990  
 Db 580 rgValAlaAsnMetAlaPheSerAsnAspLeuPheAsnLysSerSerLeuGluTyrArgA 600  
 QY 1991 CTCTGGAGCACAATTCACAGCTGTGTTCCATATCTACGATCCCAATCTTACAGAT 2050  
 Db 600 laLeuGlnGlnPheThrGlnLeuLeuValProTyrLeuArgSerAsnLeuThrGlyP 620  
 QY 2051 TTAAGCAACTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2110  
 Db 620 helysGlnLeuGluIleLeuAsnPheArgAsnGlySerValIleValAsnSerLysMetL 640  
 QY 2111 AGTTTGTAGTCTGTGGCGGTATTAACCTCAAGGCTGTGACGGGCTCTTGGAGGAT 2170



Db 380 leAlaGlySerLeuProAlaPheGlyProAspThrGlnSerGluLeuProThrSerPheA 400  
QY 1391 CTGTTATACAGAGGAGTCTACTTTGAGTCCAGAGCTTCTCTCTGTTGTAACCCAGCTTG 1450  
Db 400 laValleThrGluAAspAlaThrLeuSerProGluLeuProProValGluProGlnLeuG 420  
QY 1451 AGACAGTGCAGGAGCAGAGCATGCTTACTCAGACACTCTTGCTTCCACCTGATAGG 1510  
Db 420 luThrValAspGlyAlaGluHisGlyLeuProAspThrSerTrpSerProAlaMetA 440  
QY 1511 CCTTACCTCCCTGTCAGAGCTCCACTTCTTTATGSCATCAACACTCTCTCTGTA 1570  
Db 440 laSerThrSerLeuSerGluAlaProProPhePheWecAlaSerSerIlePheSerLeuT 460  
QY 1571 CTGATCAGGCACACAGATCAATGCCACTGACCAGACAATGCTAGTACCAGGGCTCA 1630  
Db 460 hrAspGlnGlyThrThrAspThrMetAlaThrAspGlnThrMetLeuValProGlyLeuT 480  
QY 1631 CCATCCCCACACGATGATTCTGCATCAGCCAACTGGCTCTGGGAATTCACATCCAC 1690  
Db 480 hrIleProThrSerAspTyrSerAlaIleSerGlnLeuAlaLeuGlyIleSerHisProP 500  
QY 1691 CTGCATCTTCAGATGACAGCCGATCAAGTGCAGGTGGCGAAGATATGTCAGACACCTAG 1750  
Db 500 roAlaSerSerAspAspSerArgSerSerAlaGlyGlyGluAspMetValArgHisLeuA 520  
QY 1751 ATGAATGATCTGTCGACACTCTGCTGCCCATCTGAGGTACAGAGCTCAGCGAATATG 1810  
Db 520 spGluMetAspLeuSerAspThrProAlaProSerGluValProGluLeuSerGluTyrV 540  
QY 1811 TTCTGTCTCCACATCATTTCTGAGGATACACTCTGCTCTCAGCTTTACAGTATATCA 1870  
Db 540 alSerValProAspHisPheLeuGluAspThrThrProValSerAlaLeuGlnTyrIleT 560  
QY 1871 CCACTAGTTCTATGACCAATGCCCCCAAGGCCGAGAGCTGGTAGTGTCTCAGCTCCG 1930  
Db 560 hrThrSerSerMetThrIleAlaProLysGlyArgGluLeuValValPhePheSerLeuA 580  
QY 1931 GTGTCTGTAACATGGCTTCTCCACGACCTGTGTCACAGAGCTCTCTGGAGTACCAG 1990  
Db 580 rgValAlaAsnMetAlaPheSerAsnAspLeuPheAsnLysSerSerLeuGluTyrArgA 600  
QY 1991 CTCTGGAGCAACAATTCACACAGCTGCTGCTTCATATCTACGATCCATCTTACAGGAT 2050  
Db 600 laLeuGluGlnPheThrGlnLeuLeuValProTyrLeuArgSerAsnLeuThrGlyP 620  
QY 2051 TTAAGCACTTGAATATCTTAATCTCAGAAACGGGAGTGTGATTGTGAATAGCAAAATGA 2110  
Db 620 heLysGlnLeuGluLeuAsnPheArgAsnGlySerValIleValAsnSerLysMetL 640  
QY 2111 AGTTTGCTAAGTCTGCGTATACCTTACCAAGCTGTGACGGGGTCTTGAGGATT 2170  
Db 640 ysPheAlaLysSerValProTyrAsnLeuThrLysAlaValHisGlyValLeuGluAAsp 660  
QY 2171 TTGTTTCTGTCGAGCCCAACAATCTCCATCTGGAATAGACAGCTCTCTCACAATG 2230  
Db 660 heArgSerAlaAlaAlaGlnLeuHisLeuGluLeuAspSerTyrSerLeuAnIleG 680  
QY 2231 AACAGCTGTATCAGAGCATCCCTGCAAGTCTCTGCGCTGGCGGAATTTGCCAATGTG 2290  
Db 680 luProAlaAspGlnAlaAspProCysLysPheLeuAlaCysGlyGluPheAlaGlnCysV 700  
QY 2291 TAAAGCAACAGGACTGAGGAGCGGAGTGTGCTGCGAACCAGGATATGACGCCAGG 2350  
Db 700 alLysAsnGluArgThrGluGluAlaGluCysArgCysLysProGlyTyrAspSerGlnG 720  
QY 2351 GGAGCTCTGGAGCTCTGGAAACAGGCTCTGTGCGCTGGCACAAGGATCGAGTCC 2410  
Db 720 lYserLeuAspGlyLeuGluProGlyLeuCysGlyLeuAlaGlnArgAsnAlaArgSer 739  
QY 2411 TCCAGGGAGAGGAGTCCATGCGAGGTGCGAGATCACTCTGAAAATCAAGCATCAAAA 2470  
||| :|||

Db 740 SerArgGluArgGluLeuHisAla--ValProAspHisSerGluAsnGlnAlaTyrLysT 759  
QY 2471 CTAGTGTAAAAAGTTCCAAAAATCAACAAAATAACAAGG 2509  
Db 759 hrSerVal-LysSerSerLysIleAsnLysIleThrArg 771  
RESULT 3  
US-10-007-270-4  
; Sequence 4, Application US/10007270  
; Publication No. US20020160954A1  
; GENERAL INFORMATION:  
; APPLICANT: Hageman, Gregory S.  
; APPLICANT: Kuehn, Markus H.  
; APPLICANT: University of Iowa Research Foundation  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS FOR OCULAR ABNORMALITIES  
; FILE REFERENCE: 020618-000120US  
; CURRENT APPLICATION NUMBER: US/10/007,270  
; PRIOR FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 09/430,195  
; PRIOR FILING DATE: 1999-10-29  
; PRIOR APPLICATION NUMBER: US 09/183,972  
; PRIOR FILING DATE: 1998-10-29  
; NUMBER OF SEQ ID NOS: 37  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 719  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Human IPM 150 amino acid sequence, isoform B  
US-10-007-270-4  
Alignment Scores:  
Pred. No.: 2,6e-291 Length: 719  
Score: 3611.00 Matches: 717  
Percent Similarity: 87.67% Conservative: 1  
Best Local Similarity: 87.55% Mismatches: 1  
Query Match: 61.46% Indels: 100  
DB: 13 Gaps: 2  
US-10-007-270-1 (1-3330) x US-10-007-270-4 (1-719)  
QY 131 ATGTTATTGGAACCTAGAAAGCTATTTTGTGTTTGTGATTCTTCTCAAGTTCAGCA 190  
|||  
Db 1 MetTyrLeuGluThrArgA-galallePheValPheIrrPheLeuGlnValGlnGly 20  
QY 191 ACCAAGATATCTCCATTAACATATACCATCTTGGAACTAAAGACATAGACAATCCCCCA 250  
|||  
Db 21 ThrLys----- 22  
QY 251 AGAAATGAACAACTGAAAGTACTGAAAAAATGTACAAAATGTCAACTATGAGACGAATA 310  
|||  
Db 22 ----- 22  
QY 311 TTCGATTGGCAAGCATCGAAGAAAAGATCCGCGATTTTCCCAACGGGGGTTAAAGTC 370  
|||  
Db 22 ----- 22  
QY 371 TGTCACAGGAATCCATGAAACAGATTTTAGACAGTCTTCAAGCTTATTATAGATTGAGA 430  
|||  
Db 22 ----- 22  
QY 431 GTGTGTGAGGAAGCAGTATGGAAGCATATCGGATCTTCTGGATCCATCCCTGACACA 490  
|||  
Db 23 ValCysGlnGluAlaValIrrPheLeuAlaTyrArgIlePheLeuAspArgIleProAspThr 42  
QY 491 GGGAATATCAGGATGGGTGAGCATCTGCGCAGCAGAGACTTCTCTCCCTTTTGACATT 550  
|||  
Db 43 GlyGluTyrGlnAspTrpValSerIleCysGlnGlnGluThrPheCysLeuPheAspIle 62  
QY 551 GGAATAACTTACGCAATTTCCAGAGACCTCGATCTTCTCCAGCAGAGAAATAAACAAG 610  
|||  
Db 63 GlyLysAsnPheSerAsnSerGlnGluHisLeuAspLeuLeuGlnArgIleLysGln 82

QY 611 AGAAGTTCCCTGACAGAAAGATGAATATCTGCAGAGAGAGATTTGGAGAGCCCTGGT 670  
 Db 83 ArgSerPheProAspArgLysAspGluLeuSerAlaGluLeuThrLeuGluProGly 102  
 QY 671 GAAACCATTTGATCTCAACAGCAATCTACATTTCAAAGACTTGGGCGAGTATCTAAGAA 730  
 Db 103 GluThrIleValIleSer--Thr----- 109  
 QY 731 AACCCCTCAGAGACCAAAATTCAGAGTGTGCAACAGTCTCACTTGGGCGCTTCCCTCTCA 790  
 Db 110 -----AspValAlaAsnValSerLeuGlyProPheProLeu 122  
 QY 791 CTCCTGATGACACCCCTCCTCAATGAAATCTCGATTAATACATCAACAGCAACCAAGATGC 850  
 Db 122 hrProAspThrLeuLeuAsnGluLeuAspAsnThrLeuAsnAspThrLysMetp 142  
 QY 851 CTACACAGAAAGAAACAGAATTGCTGTGTGGAGGAGAGAGGCTGAGCTCAGCTCAGCTCAGG 910  
 Db 142 roThrGluArgGluThrGluPheAlaValLeuGluGluGlnArgValGluLeuSerV 162  
 QY 911 TCTCTCTGTTAAACCAAGATTCAGAGTGTGCTGATCCAGCTCAGCTCAGCTCAGCTCAGCT 970  
 Db 162 alSerLeuValAsnGlnLysPheLysAlaGluLeuAlaAspSerGlnSerProTyrTyrG 182  
 QY 971 AGGAGTAGCAGGAAAGTCCCACTTCAGATGCAAAAGATATTTAAGAACTTCAGAT 1030  
 Db 182 InGluLeuAlaGlyLysSerGlnLeuGlnMetGlnLysIlePheLysLysLeuProGlyp 202  
 QY 1031 TCMAAAATTCATGTGTAGATTTAGACCAAGAAAGAAAGATGCTCAGCTCAGCTCAGCT 1090  
 Db 202 helyLysIleHisValLeuGlyPheArgProLysLysGluLysAspGlySerSerSert 222  
 QY 1091 CAGAGTGCACACTTACGCCATCTTTAAGAGACACAGTGCAGAGCAAAAGCCCTGCAA 1150  
 Db 222 hrGluMetGlnLeuThrAlaIlePheLysArgHisSerAlaGluAlaLysSerProAlas 242  
 QY 1151 GTGACCTCTGCTTTTGAATTCACAAATTTGAAGTGAAGAGTCTATCATGGAACCA 1210  
 Db 242 erAspLeuLeuSerPheAspSerAsnLysIleGluSerGluGluValTyrHisGlyThrM 262  
 QY 1211 TGAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1270  
 Db 262 etGluGluAspLysGlnProGluIleTyrLeuThrAlaThrAspLeuLysArgLeuLys 282  
 QY 1271 GCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1330  
 Db 282 erLysAlaLeuGluGluGlnSerLeuAspValGlyThrIleGlnPheThrAspGluI 302  
 QY 1331 TTCTGGATCACTGCCAGCCTTTGGTCTGACACCCCAATCAGAGCTGCCACATCTTTTG 1390  
 Db 302 leAlaGlySerLeuProAlaPheGlyProAspThrGlnSerGluLeuProThrSerPheA 322  
 QY 1391 CTCTTATACAGAGATGCTTCTGAGTCCAGACTTCTCTCTGTTGAACCCAGCTTG 1450  
 Db 322 laValIleThrGluAspAlaThrLeuSerProGluLeuProValGluProGlnLeuG 342  
 QY 1451 AGACAGTGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1510  
 Db 342 luThrValAspGlyAlaGluHisGlyLeuProAspThrSerTyrProAlaMetA 362  
 QY 1511 CCTCTACCTCTCTGTCAGAGCTCCACTTTCTTTATGTCATCAAGCATCTCTCTCTGA 1570  
 Db 362 laSerThrSerLeuSerGluAlaProProPhePheMetAlaSerSerIlePheSerLeu 382  
 QY 1571 CTGATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1630  
 Db 382 hrAspGlnGlyThrThrAspThrMetAlaThrAspGlnThrMetLeuValProGlyLeu 402  
 QY 1631 CCATCCCCCAGCAGTATTCTGCATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1690  
 Db 402 hrIleProThrSerAspTyrSerAlaIleSerGlnLeuAlaLeuGlyIleSerHisProp 422

QY 1691 CTGCATCTTTCAGATGACAGCCGATCAAGTTCAGGTGGCGAAGATATGTCAGACACTAG 1750  
 Db 422 roAlaSerSerAspAspArgSerSerAlaGlyGlyGluAspMetValArgHisLeuA 442  
 QY 1751 ATGAAATGGAATCTGCTCAGACCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1810  
 Db 442 spGluMetAspLeuSerAspThrProAlaProSerGluValProGlyLeuSerGluTyrV 462  
 QY 1811 TTTCTGTCACGATCATTTCTTGAGAGTACCACTCTCTGCTCAGCTTACAGTATATCA 1870  
 Db 462 alSerValProAspHisPheLeuGluAspThrProValSerAlaLeuGlnTyrIleT 482  
 QY 1871 CCCTAGTTCATGACCATTTGCCCCCAAGGCCGAGAGCTGTGTAGTGTCTTTCAGTCTGC 1930  
 Db 482 hrThrSerSerMetThrIleAlaProLysGlyArgGluLeuValValPhePheSerLeuA 502  
 QY 1931 GTGTTCTTAAACATGCTCTTCCACGACCTGTTCAACAGAGCTCTCTGAGTACCGAG 1990  
 Db 502 rgValAlaAsnMetAlaPheSerAsnAspLeuPheAsnLysSerSerLeuGluTyrArgA 522  
 QY 1991 CTCTGAGGCAATTCACACAGCTCTGCTGCTTCCATATCTAGATCCCAATCTTACAGAT 2050  
 Db 522 laLeuGluGlnPheThrGlnLeuLeuValProTyrLeuArgSerAsnLeuThrGlyp 542  
 QY 2051 TTAAGCAACTTGAATACTTAATTCAGAACCGGAGTGTGATTGTGAATAGCAAAATGA 2110  
 Db 542 helyGlnLeuGluIleLeuAsnPheArgAsnGlySerValIleValAsnSerLysMetL 562  
 QY 2111 AGTTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2170  
 Db 562 yspPheAlaLysSerValProTyrAsnLeuThrLysAlaValHisGlyValLeuGluAsp 582  
 QY 2171 TTTGCTTCTGCTGAGCCCAACAACTCCATCTGGAATAGACAGCTACTCTCTCAACATTG 2230  
 Db 582 heArgSerAlaAlaAlaGlnGlnLeuHisLeuGluIleAspSerTyrSerLeuAsnIleG 602  
 QY 2231 AACAGCTGATCAAGCAGATCCCTGCAAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2290  
 Db 602 luProAlaAspGlnAlaAspProCysLysPheLeuAlaCysGlyGluPheAlaGlnCysV 622  
 QY 2291 TAAAGAACGACGACTGAGAAAGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2350  
 Db 622 alLysAsnGluArgThrGluGluAlaGluCysArgCysLysProGlyTyrAspSerGlnG 642  
 QY 2351 GAGCCTTGACCGCTCTGACACAGGCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2410  
 Db 642 lysSerLeuAspGlyLeuGluProGlyLeuCysGlyProGlyThrLysGluCysGluVal 662  
 QY 2411 TCAGAGGAAAGCGAGCTCCATGACAGTTCGACAGTTCGACAGTTCGACAGTTCGACAGTTCGACAG 2470  
 Db 662 euGlnGlyLysGlyAlaProCysArgLeuProAspHisSerGluAsnGlnAlaTyrLysP 682  
 QY 2471 CTAGTCTTAAAGTTCCTCAAAATCAACAAATTAACAGGTAAATCAAGTAAAGAAATCTG 2530  
 Db 682 hrSerValLysLysPheGlnAsnGlnGlnAsnLysValIleSerLysArgAsnSerG 702  
 QY 2531 AATTACTGACCTGATATATGAAATTAACATCAAGATTTACCATCAAGATTTGGAAGAAAT 2583  
 Db 702 luLeuLeuThrValGluTyrGluGluPheAsnHisGlnAspTrpGluGlyAsn 719

RESULT 4  
 US-10-007-270-9  
 ; Sequence 9, Application US/10007270  
 ; Publication No. US20020160954A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hageman, Gregory S.  
 ; APPLICANT: Kuehn, Markus H.  
 ; APPLICANT: University of Iowa Research Foundation  
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS FOR OCULAR ABNORMALITIES  
 ; FILE REFERENCE: 020618-00012005  
 ; CURRENT APPLICATION NUMBER: US/10/007,270  
 ; PRIOR FILING DATE: 2001-11-08  
 ; PRIOR APPLICATION NUMBER: US 09/430,195

; PRIOR FILING DATE: 1999-10-29  
 ; PRIOR APPLICATION NUMBER: US 09/183,972  
 ; NUMBER OF SEQ. IDS: 37  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ. ID NO. 9  
 ; LENGTH: 798  
 ; TYPE: PRT  
 ; ORGANISM: Mus sp.  
 ; FEATURE:  
 ; OTHER INFORMATION: Mouse IPM 150 amino acid sequence, isoform A  
 US-10-007-270-9

# Alignment Scores:

Pred. No.: 3,17e-199 Length: 798  
 Score: 2504.50  
 Percent Similarity: 73.13% Matches: 527  
 Best Local Similarity: 80 Conservative: 80  
 Query Match: 63.49% Mismatches: 180  
 DB: 42.63% Indels: 43  
 13 Gaps: 10

US-10-007-270-1 (1-3330) x US-10-007-270-9 (1-798)

|    |     |  |     |
|----|-----|--|-----|
| QY | 131 | ATGCTATTGGAACTAGAGAGCTATTTTGTGTTTTCCTCAAGTTCAGGA             | 190 |
| Db | 1   | MetAsnPheGlnIleLysHisAlaIlePheValPheGlyIlePheLeuGlnValGlnGly | 20  |
| QY | 191 | ACCAAGATATTCCTCAATTAACATATACCATCTCTGAACTTAAGACATAGACATCC     | 250 |
| Db | 21  | IleLysAspThrSerIleLysIlePheSerSerGluIleLysAsnIleAspLysThrPro | 40  |
| QY | 251 | AGAAATGAACAACTGAAAGTACTGAAATAATGTACAAATGTCAACTATGAGCAATA     | 310 |
| Db | 41  | ArgIleGluThrIleGluSerThrSerValHisLysValSerThrMetIleArgIle    | 60  |
| QY | 311 | TTCCGATTGGCAAGCATCGAAACAAAGATCCGATTTTCCCAACGGGGGTTAAAGTC     | 370 |
| Db | 61  | PheAspLeuProLysLeuArgThrLysArgSerAlaLeuPhePro---AlaAlaAsnIle | 79  |
| QY | 371 | TGTCACAGGATTCATCAACAGATTTTAGACAGTCTTCAAGCTTTATATAGATTGAGA    | 430 |
| Db | 80  | CysProGlnIleSerLeuArgGlnIleLeuAlaSerLeuGlnIleLysArgLeuArg    | 99  |
| QY | 431 | GTGTGTGACAGACAGTATGGAGCATATCGATCTTCTGATCGATCCGATCCGACACA     | 490 |
| Db | 100 | ValCysGlnGluValIleProGluAlaTyrArgIlePheLeuAspArgIleProAspThr | 119 |
| QY | 491 | GGGGAATATCAGACTGGGTGACATCTCCAGAGGACCTTCTGCTCTTTGACATT        | 550 |
| Db | 120 | GluGluTyrGlnAspTrpValSerLeuCysGlnLysGluThrPheCysLeuPheAspIle | 139 |
| QY | 551 | GGMAAAATTCAGCAATTCAGGAGACCTGATCTTCTCCAGCAGATATAAAGCAG        | 610 |
| Db | 140 | GlyLysAsnPheSerAsnSerGlnGluHisLeuAspLeuGlnGlnArgIleLysGln    | 159 |
| QY | 611 | AGAAGTTTCCCTGACAGAAAGATGAATATCTGACAGAGACATTTGGAGAGCCTGT      | 670 |
| Db | 160 | ArgSerPheProGlyArgLysAspGluThrAlaSerMetGluThrLeuGluAlaProThr | 179 |
| QY | 671 | GAACACATTTGATTTCAACAGCATCTACATTTTCAAGACTTGGGAGATTTCTAAGAA    | 730 |
| Db | 180 | GluAlaProValPro---Thr-----                                   | 186 |
| QY | 731 | AACCTTCAGAGAGCAAAATTCAGATGTTGCCAGCTCTCACTTGGGCTTCCCTCTCA     | 790 |
| Db | 187 | -----AspValSerArgMetSerLeuGlyProPheProLeuP                   | 199 |
| QY | 791 | CTCTGTATGACACCTCTCAATGAATTTCTCGATTAATCACTCAACACACCAAGATGC    | 850 |
| Db | 199 | roSerAspAspThrAspLeuLysGluIleLeuSerValThrLeuLysAspIleGlnLysP | 219 |
| QY | 851 | CTACACAGAGAGAAACA-----GAATTCGCTGTGTTGGAGGAGC                 | 892 |
|    |     |  |     |

|    |      |  |      |
|----|------|--|------|
| Db | 219  | roThrThrGluSerLysThrGluProIleHisValSerGluPheSer-----SerGluG  | 237  |
| QY | 893  | AGAGGTGGAGCTCAGCGTCTCTGTAAACCAAGAAAGTTCAAGGAGAGCTCGCTGACT    | 952  |
| Db | 237  | IuLysValGluPheSerIleSerLeuProAsnHisArgPheLysAlaGluLeuThrAsnS | 257  |
| QY | 953  | CCAGTCCCATATACAGGAGCTAGCAGCAAGATCCCACTTCAGATGCAAAAGATAT      | 1012 |
| Db | 257  | erGlySerProTyrTyrGlnGluLeuValGlyGlnSerGlnLeuGlnLysIleP       | 277  |
| QY | 1013 | TTAAGAAATCCAGGATTCAAAATAATCCATGTGTAGATTAGACCAAGAAAGAA        | 1072 |
| Db | 277  | helyLysLeuProGlyPheGlyGluIleArgValLeuGlyPheArgProLysGluG     | 297  |
| QY | 1073 | AAGATGCTCAAGCTCCAGAGATGCAACTTAGGCCATCTTTAAGAGACACAGTGCGAG    | 1132 |
| Db | 297  | IuAspGlySerSerThrGluIleGlnLeuMetAlaIlePheLysArgAspHisAlaG    | 317  |
| QY | 1133 | AAGCAAAAGCCCTGCAAGTACCTCTGCTGTTTTCATTCACCAAAATTTGAAGTGAGG    | 1192 |
| Db | 317  | IuAlaLysSerProAspSerHisLeuLeuSerLeuAspSerAsnLysIleGluSerGlu  | 337  |
| QY | 1193 | AGTCTATCTGGAACCAATGAGGAGGAGCAAGCAACCAAAATCTATCTCAGAGCTACAG   | 1252 |
| Db | 337  | rgIleHisIleGlyValIle---GluAspLysGlnProGluThrTyrLeuThrAlaThr  | 356  |
| QY | 1253 | ACCTCAAAAGGCTGATCACAAGCATAGAGGAGCAACATCTTGGATGGGGCAAA        | 1312 |
| Db | 356  | spLysLysLeuIleGlnLeuLeuAspGlyAspLeuSerLeuValGluGlyLysI       | 376  |
| QY | 1313 | TTCACTTCACTGATGAATGCTGATCACTGCCAGCTTGTGGCTGACACCCATCAG       | 1372 |
| Db | 376  | leProPheGlyAspGluValThrGlyThrLeu-----PheArgProValThrGluPro   | 394  |
| QY | 1373 | AGCTCCCAACATCTTTGCTGTATTAACAGAGATGCTACTTTGAGTCAGCAACTCTCTC   | 1432 |
| Db | 394  | spLeuProLysProLeuAlaAspValThrGluAspAlaThrLeuSerProGluLeuPro  | 414  |
| QY | 1433 | CTGTGTAAACCCAGCTCAGACAGTGGAGGAGGAGAGCATGGTCTACCT-----        | 1482 |
| Db | 414  | heValGluProArgLeuGluAlaValAspArgGluGlySerGluLeuProGlyMetSerS | 434  |
| QY | 1483 | -----GACACTTTGCTCCACCTGCTATGCTCTTACCTCTCTCTCAGAGCTCCAC       | 1537 |
| Db | 434  | erLysAspSerSerTrpSerProValSerAlaSerIleSerArgSerGluAsnLeuP    | 454  |
| QY | 1538 | CTTTCTTATGGATCAAGCATCTCTCTGATCAAGGAGGAGGAGGAGGAGGAGGAGG      | 1597 |
| Db | 454  | roSerPhe---ThrProSerIlePheSerLeuAspAlaGlnSerProProLeuMetT    | 473  |
| QY | 1598 | CCACTGACCAACAATGCTAGTACAGGGCTCACCATCCCAAGGAGGAGGAGGAGGAGG    | 1657 |
| Db | 473  | hrThrGlyProThrAlaLeuIleProLysProThrLeuProThrIleAspTyrSerThrI | 493  |
| QY | 1658 | TCAGCAACTGCTCTGGGAATTTCACTCCACCTGATCTTCAGATCAGACCGCATCA      | 1717 |
| Db | 493  | leArgGlnLeuProLeuGluSerSerHisTrpProLaserSerSerAspArgGluLeuI  | 513  |
| QY | 1718 | GTCAGGTGGCAAGATATGTCAGACACCTAGATGAATGGATCTGTCTGACACTCTCTG    | 1777 |
| Db | 513  | leThrSerSerHisAspThrIleArgAspLeuAspGlyMetAspValSerAspThrPro  | 533  |
| QY | 1778 | CCCATCTGAGTACAGAGCTCAGCAATATGTTCTCTCCAGATCATTTCTTGAGG        | 1837 |
| Db | 533  | IaLeuSerGluIleSerGluLeuSerGlyTyrAspSerAlaSerGlyGlnPheLeuGlu  | 553  |
| QY | 1838 | ATACACTCTGCTCTAGCTTTACAGTATATCACCTAGTGTCTTCTACCATTTGCCCA     | 1897 |
| Db | 553  | etThrThrProIleProThrValArgPheIleThrSerSerGluThrIleAlaThrL    | 573  |
| QY | 1898 | AGGCGGAGAGCTGTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG    | 1957 |
| Db | 573  | ysGlyGlnGluLeuValPhePheSerLeuArgValAlaAsnMetProPheSerTyrA    | 593  |

```

1958 ACCTGTTCAACAGAGCTCTCTGGAGTACCGAGCTCTGGAGCAACAATTCAACAGCTGC 2017
Db spLeuPheAsnLysSerSerLeuGlnValGlnAlaLeuGlnArgPheThrAspLeuL 613
2018 TGGTTCATATCTACGATCCAACTTTACAGGATTTAAGCAACTGGAATATTAACCTCA 2077
Db euValProTyrLeuArgSerAsnLeuThrGlyPheLysGlnLeuGluLeuSerPheA 633
2078 GAAACGGAGTGTGATCTGTAATAGCAAAATGAAGTTGCTGAAGTCTGTGCGGTATAACC 2137
Db rglanGlySerValIleValAsnSerLysValArgPheAlaLysAlaValProTyrAsnL 653
2138 TCACCAAGCTGTGCACGGGCTCTCGAGGATTTTCGTTCTGCTGCACCCCAACAACCTCC 2197
Db euThrGlnAlaValArgGlyValLeuGluValAspLeuArgSerThrAlaAlaGlnGlyLeuA 673
2198 ATCTGGAATAGACAGCTACTCTCAACATTGAACAGCTGATCAAGCAGATCCCTGCA 2257
Db rglanGlySerValIleValAsnSerLysValArgPheAlaLysAlaValProTyrAsnL 653
2258 AGTTCCTGCTGCGCTGGCGCAATTTGCCAATGTGTAAAGCAAGCGAGTGAAGCGG 2317
Db yslLeuLeuAspCysGlyLysPheAlaGlnCysValLysAsnGluTyrThrGluAlaIag 713
2318 AGTTCGCTGCAACAGCAGATATGACAGCCAGGAGGAGCTGACGCTGCGTGGAAACAGGCC 2377
Db luCysArgCysArgGlnGlyHisGluSerHisGlyThrLeuAspTyrGlnThrLeuAsnL 733
2378 TCTGTGGCCCTGGGCAAAAGGAATCGAGGCTCTCCAGGAAAGGAGGAGCTCCATCAGGT 2437
Db euCysProProGly--LysThrCysValAlaGlyArgGluGlnAlaThrProCysArgP 752
2438 TGCCAGATCACTCTGAAATCAACGATCAAACTAGTGTAAAGTTTCAAAATCAAC 2497
Db roProAspHisSerThrAsnGlnAlaGlnGluProGlyValLysLysLeu-----ArgG 770
2498 AAAATAACAAGTAACTACAGTAAAGAAATTTCTGAATTAAGTACCGCTAGATATGAAGAT 2557
Db lngInAsnLysValLysLysArgAsnSerLysLeuSerAlaIleGlyPheGluLup 790
2558 TTAACCATCAAGATTGGGAAGAAAT 2583
Db heGluAspGlnAspTrpGluGlyAsn 798

RESULT 5
US-10-007-270-11
; Sequence 11, Application US/10007270
; Publication No. US20020160954A1
; GENERAL INFORMATION:
; APPLICANT: Hageman, Gregory S.
; APPLICANT: Kuehn, Markus H.
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS FOR OCULAR ABNORMALITIES
; FILE REFERENCE: 020618-000120US
; CURRENT APPLICATION NUMBER: US/10/007,270
; PRIOR FILING DATE: 2001-11-08
; PRIOR FILING DATE: 1999-10-29
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Mus sp.
; FEATURE:
; OTHER INFORMATION: Mouse 1P50 amino acid sequence, isoform D
US-10-007-270-11
Alignment Scores:
Pred. No.: 5.56e-99 Length: 466

```

```

Score: 1298.50 Matches: 309
Percent Similarity: 43.76% Conservative: 52
Best Local Similarity: 37.45% Mismatches: 99
Query Match: 22.10% Indels: 365
DB: 13 Gaps: 10

US-10-007-270-1 (1-3330) x US-10-007-270-11 (1-4666)

QY 131 ATGTATTGGAACTAGAGAGCTATTATTTTGGATTTTTCCTCAAGTTCAAGGA 190
Db 1 MetAsnPheGlnIleLysHisAlaIlePheValPheGlyIlePheLeuGlnValGlnGly 20
QY 191 ACCAAAGATATTCCTAATTAAACATATACCATCTCTGAAACTAAAGACATAGACATCCCCCA 250
Db 20 ----- 20
QY 251 AGAAATCAAAACAACCTGAAGTACTGAAATAATGTACAAATGTCAAAATGTCAAGTATGAGACGAATA 310
Db 20 ----- 20
QY 311 TTCGATTGGCAAGCATCGAACCAAAAGATCCGCATTTTCCCACCGGGGTAAAGTC 370
Db 20 ----- 20
QY 371 TGTCCACAGGAATCCATGAAACAGATTTTACACAGTCTTCAAGCTTATTATAGATTGACA 430
Db 21 ----- 11Lys 22
QY 431 GTGTGTGAGGAAGCAGTATGGGAAGCATATCGAATCTTTCGTGATCGCATCCCTGACACA 490
Db 23 ValCysGlnGluValValTrpGluAlaTyrArgIlePheLeuAspArgIleProAspThr 42
QY 491 GGGGAATATCAGGACTGGGTCAAGCATCTGCAGCAGCAGACCTTCCTCTCTTTCGATTT 550
Db 43 GluGluTyrGlnAspTrpValSerLeuGlnLysGluThrPheCysLeuPheAspIle 62
QY 551 GGAATAAACTTCAGCAATTCCTCCAGAGCAGCCTGGATCTTCTCCAGCAGAGATAAAGCAG 610
Db 63 GlyLysAsnPheSerAsnSerGlnGluHisLeuAspLeuGlnGlnArgIleLysGln 82
QY 611 AGAGTTTCCCTCAGCAAAAGATGAAATATCTGCAGAGAGAGCATTTGGGAGAGCTGGT 670
Db 83 ArgSerPheProGlyArgLysAspGluThrAlaSerMetGluThrLeuGluAlaProThr 102
QY 671 GAAACCATTTCTCATTTCAACAGCAATCTACATTTCAAAGACTTGGGAGATTTCTAAGAA 730
Db 103 GluAlaProValPro--Thr----- 109
QY 731 AACCTCAGAGAGCAAAATTCAGATGTGGCAAGTCTCACTTGGGCTTTTCCCTCTCA 790
Db 110 ----- 122
QY 791 CTCCTGATGACACCTCTCTCAATGAAATTCGATATATACACTCAACAGACCAAGATGC 850
Db 122 roSerAspThrAspLeuLysGluIleLeuSerValThrLeuLysAspIleGlnLysP 142
QY 851 CTACACACAAAGAGAAACA-----GAAATTCGCTGTGTGGAGGAGC 892
Db 142 roThrGluSerIleThrGluProIleHisValSerGluPheSer-----SerGluG 160
QY 893 AGAGGTGGAGCTCAGGCTCTCTCTGGTAAACCAAGAGTTCAAGCAGAGCTCGCTGACT 952
Db 160 LysValGluPheSerIleSerLeuProAsnHisArgPheLysAlaGluLeuThrAsnS 180
QY 953 CCAGTCCCATATATTACAGGAGCTAGCAGGAAAGTCCCACTTCAGATCAAGAAATAT 1012
Db 180 erGlySerProTyrTyrGlnGluLeuValGlyGlnSerGlnLeuGlnLysIleP 200
QY 1013 TTAAGAAATTCGAGATTTCAAAAAATTCATGTGTAGGATTTTAGACCAAGAGAGAA 1072
Db 200 helLysLeuProGlyPheGlyGluIleArgValLeuGlyPheArgProLysGluG 220
QY 1073 AAGATGGCTCAAGCTCCACAGAGATGCAACTTACGGCCACTTTTAAAGACACAGTGCAG 1132

```



```
Db 220 luAspGlySerSerThrGluLeuMetAlaIlePheLysArgAspHisAlaG 240
:::|||||
QY 1133 AAGCAAAAGCCCTCGAAGTGAAGCTCTCTCTTTTGGATTCCAAACAAATTTGAAAGTGAGG 1192
|||||
Db 240 luAlaLysSerProAspSerHisLeuLeuSerLeuAspSerAsnLysIleGluSerGluA 260
|||||
QY 1193 AAGTCTATCATGGAACCATGAGGAGGAGCAAGCAACCAAGCAATCTATCTTCACAGCTACAG 1252
|||||
Db 260 rgIleHisGlyValIle--GluAspLysGlnProGluThrTyrLeuThrAlaThra 279
:::|||||
QY 1253 ACCTCAAAAGCGTGAATGCAAGCAAGCACTAGAGGAAGCAAACTTTGGATGTGGGACAA 1312
|||||
Db 279 spLeuLysLysLeuIleLeuLeuLeuAspGlyAspLeuSerLeuValGluGlyLysI 299
|||||
QY 1313 TTCAAGTTCACATGCAAAATGCTGGATCACTGCCAGCTTTGGTCTTGACACCAATCAG 1372
|||||
Db 299 leProPheLysGluValThrGlyThrLeu-----PheArgProValThrGluProA 317
|||||
QY 1373 AGCTGCCACATCTTTTGTCTGTATAACAGAGGATGCTACTTTGAGTCCAGAACTTCTC 1432
:::|||||
Db 317 spLeuProLysProLeuAlaAspValThrGluAspAlaThrLeuSerProGluLeuProp 337
|||||
QY 1433 CTGTTGAACCCACCTTGACAGCTGGAGCGGAGAGCATGGTCTACCTGACACTTCTT 1492
|||||
Db 337 heValGluProArgLeuGluAlaValAsp----- 346
|||||
QY 1493 GGTCTCCACCTGTATGGCTCTACTCTCTGTGTCAGAGCTCCACCTTTCTTTATGGCAT 1552
|||||
Db 346 ----- 346
QY 1553 CAAGCATCTTCTCTGACTGATCAAGGCACACAGATACAATGGCCACTGCACAGACAA 1612
|||||
Db 346 ----- 346
QY 1613 TGCTAGTACAGGGCTCACCATCCCAACCAAGTATTTCTGCAATCAGCAACTGGCTC 1672
|||||
Db 346 ----- 346
QY 1673 TGGGAATTCACATCCACCTGTCATCTTCAGATGACAGCGGATCAAGTGCAGGTGGCGAAG 1732
|||||
Db 346 ----- 346
QY 1733 ATATGTGTCAGACACCTAGATGAATGATCTGTGTGACACTCTGCTGCCCATCTGAGGTAC 1792
|||||
Db 346 ----- 346
QY 1793 CAGAGCTCAGCGAATATGTTCTGTCTCCAGATCATTTCTTGGAGGATACCACTCTGTCT 1852
|||||
Db 346 ----- 346
QY 1853 CAGCTTTACAGTATATACCACTAGTCTTATGACCATTTGCCCAAGGCGCGAGAGTGG 1912
|||||
Db 346 ----- 346
QY 1913 TAGTGTTCTCAGTCTGGTGTGCTTAAATGGGCTTCTCCAAAGCAGCTGTTCACAAAGA 1972
|||||
Db 346 ----- 346
QY 1973 GCTCTCTGAGTACCGAGCTCTGGAGCAACAATTCAACAGTGTGTGTTCCATATCTAC 2032
|||||
Db 346 ----- 346
QY 2033 GATCCAATCTTACAGGATTTAAGCAACTTGAATACTTAACCTTCAGAAAGGAGGTGA 2092
|||||
Db 347 ----- 347
QY 2093 TTGTGATAGCAAAATGAAGTTTGTGAAGTCTGTGCGGTATAACCTCACCAAGGCTGTGC 2152
:::
Db 352 eu----- 352
QY 2153 ACGGGGCTTGTGAGGATTTTCGTTCTGTGTGCGCCCAACAACCTCCATCTCGAAATAGACA 2212
|||||
```

```
Db 352 ----- 352
QY 2213 GCTACTCTCTCAACATTGACAGCTGATCAAGCAGATCCCTGCAAGTTCTCTGGCCTGG 2272
|||||
Db 353 -----ProAlaAspGlnAlaAspProCysLysLeuLeuAspCysG 366
|||||
QY 2273 GCGAATTTGCCAATGCTCTAAAGAAAGCAAGCAGCTGAGGAGCGGAGTGTGCTGCAAC 2332
|||||
Db 366 lyLysPheAlaGlnCysValLysAsnGlnTrpThrGluGluAlaGluCysArgCysArg 386
|||||
QY 2333 CAGATATGACAGCCAGGAGGAGCTGACGCTCTGGAACACGAGCTCTGTGCTGCTGCA 2392
|||||
Db 386 lnGlyHisGluSerHisGlyThrLeuAspTyrGlnThrLeuAsnLeuCysProProGly 405
|||||
QY 2393 CAAAGGAATGCGAGTCTCCAGGAAAGGAGCTCCATGAGTGTGCGAGTCCAGATCATCTG 2452
|||||
Db 406 --lysThrCysValAlaGlyArgGluGlnAlaThrProCysArgProThrAspHisSerT 425
|||||
QY 2453 AAATCAAGCATACAAAACTAGTGTATAAAAGTTCCAAAATCAACAAATAACAGGTAA 2512
|||||
Db 425 hrAsnGlnAlaGlnProGlyValLysLysLeu-----ArgGlnGlnAsnLysVal 443
|||||
QY 2513 TCAGTAAAGAAATTCGAATTTACGACCGTAGAATATGAAGATTTAACCATCAAGATT 2572
|||||
Db 443 alLysLysArgAsnSerLysLeuSerAlaIleGlyPheGluGluPheGluAspGlnAspT 463
|||||
QY 2573 GCGAAGCAAAAT 2583
|||||
Db 463 rpGluGlyAsn 466
|||||
```

## RESULT 6

```
US-10-007-270-6
; Sequence 6, Application US/10007270
; Publication No. US20020160954A1
; GENERAL INFORMATION:
; APPLICANT: Hageman, Gregory S.
; APPLICANT: Kuehn, Markus H.
; APPLICANT: University of Iowa Research Foundation
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS FOR OCULAR ABNORMALITIES
; FILE REFERENCE: 020618-000120US
; CURRENT APPLICATION NUMBER: US/10/007,270
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 09/430,195
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 09/183,972
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 6
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Human IPM 150 amino acid sequence, isoform C
US-10-007-270-6
```

## Alignment Scores:

|                        |          |               |     |
|------------------------|----------|---------------|-----|
| Pred. No.:             | 1,81e-77 | Length:       | 198 |
| Score:                 | 1038.00  | Matches:      | 197 |
| Percent Similarity:    | 99.49%   | Conservative: | 0   |
| Best Local Similarity: | 99.49%   | Mismatches:   | 1   |
| Query Match:           | 17.67%   | Indels:       | 0   |
| DB:                    | 13       | Gaps:         | 0   |

US-10-007-270-1 (1-3330) x US-10-007-270-6 (1-198)

```
QY 131 ATGTATTTGGAACCTAGAGAGCTATTTTCTTTTGGATTTTCTTCAAGTTCAAGCA 190
Db 1 MetTyrLeuGluThrArgArgAlaIlePheValPheTrpIlePheLeuGlnValGlnGly 20
QY 191 ACCAAAGATATCTCATTTACATATACCATTTCTGAACTAAAGACATAGACAATCCCCA 250
Db 21 ThrLysAspIleSerIleAsnIleTyrHisSerGluThrLysAspIleAsnProPro 40
|||||
```



|    |      |  |      |
|----|------|--|------|
| Db | 357  | yrlleAlaGluThrLeuGlnGlnAsnPheLeuLeuGlyAsnSerSerLeuAsnProaspP | 377  |
| QY | 1288 | .....      ..... GAACAATCTTGGATCG-                           | 1305 |
| Db | 377  | roAspSerLeuGlnLeuIleAsnValAsgGlyValleuArgHisGlnThrGluAspLeuV | 397  |
| QY | 1306 | -----GGGACAAATCAAGTTCACT                                     | 1323 |
| Db | 397  | aLrTpAnThrGlnSerSerSerLeuGlnAlaThrProSerSerIleLeuAspAsnThrp  | 417  |
| QY | 1324 | -----GATGAATTCCTGGATCAGTGCCA                                 | 1347 |
| Db | 417  | heGlnAlaLaIrTpProSerAlaAspGluSerIleThrSerSerIleProLeuAspP    | 437  |
| QY | 1348 | -----GCCTTTGTCTCTGACCAACCATCAGAGCTGCC                        | 1380 |
| Db | 437  | heSerSerGlyProProSerAlaThrGlyArgGluLeuTrpSerGluSerProleuGlyA | 457  |
| QY | 1380 | -----  | 1380 |
| Db | 457  | spleuValSerThrHislysLeuAlaPheProSerLysMetGlyLeuSerSerSerProg | 477  |
| QY | 1381 | -----ACAT  | 1384 |
| Db | 477  | luValLeuGluValSerSerLeuThrLeuHiserValThrProAlaValLeuGlnThrG  | 497  |
| QY | 1385 | CTTTTCTCTATTATAACAGAGGATGACTATTGAGTCCAGAACATCTCTCTTTGAACCCC  | 1444 |
| Db | 497  | lyLeuProValAlaserGluGluArgThrSerGlySerHisLeu                 | 515  |
| QY | 1445 | AGCTTCAGACAGTGGAGGAGCACAG                                    | 1489 |
| Db | 515  | lyLeuAlaasnValGluGluSerGluAspPheLeuSerIleAspSerLeuProSerS    | 535  |
| QY | 1490 | CTGGTGCTCCACCTGCTATGGCTCTACC                                 | 1522 |
| Db | 535  | erPheThrGlnProValProLysGluThrIleProSerMetGluAspSerAspValSerL | 555  |
| QY | 1523 | TGTCAGAGTCCCA  | 1564 |
| Db | 555  | euthrSerSerProTyLeuThrSerSerIleProPheGlyLeuAspSerLeuThrSerL  | 575  |
| QY | 1565 | CTCTGACTGATCAAGCACCCAGATACATGGCACTGACACAGACAATG              | 1614 |
| Db | 575  | ysVallyAspGlnLeuLysValSerProPheLeuProAspAlaSerMetGluLysGluL  | 595  |
| QY | 1615 | -----CTAGTACCGGCTCAACATCC                                    | 1636 |
| Db | 595  | eullePheAspGlyGlyLeuGlySerGlyGlnLysValAspLeuIleThrTrpp       | 615  |
| QY | 1637 | CMACCAAGTATTATCTGCATCAAGCAACATGGCTCTGGGAATTTCACTACCACTGCAT   | 1696 |
| Db | 615  | rofTrpSerGluThrSerSer--GluLysSerAlaGluProLeuSerLysProTrpLeuG | 634  |
| QY | 1697 | CITTCAGATCAGCCCATCATAGTCAGGTGGGAGATATGCTCAGACACTTA---GATG    | 1753 |
| Db | 634  | luAspAspSerLeuLeuProAlacluilieGluAspLysLysLeuValLeuValAspL   | 654  |
| QY | 1754 | AAATGGATCTGCTGAC   | 1771 |
| Db | 654  | ysMetAspSerThrAspGlnileSerLysHisSerLysTyrgluHiAspAspArgserT  | 674  |
| QY | 1772 | CTCCTGCCCATCTGAGGTACCA                                       | 1794 |
| Db | 674  | hrHisPheProGluGluGluProLeuSerGlyProAlaValProIlePheAlaAspThrA | 694  |
| QY | 1795 | --GAGCTCAGCGAATATGTTCTCTCCGAGATCATTTCTTGGAG                  | 1836 |
| Db | 694  | laAlaclSerAlaSerLeuThrLeuProLysHisIleSerGluValProGlyValAspA  | 714  |
| QY | 1837 | -----GATACCATCTGTCTGAGCT                                     | 1857 |

|    |      |  |      |
|----|------|--|------|
| Db | 714  | spCysSerValThrLysAlaProLeuIleLeuThrSerValAlaIleSerAlaSerThrA | 734  |
| QY | 1858 | -----TTACAGTATATCACCACCTAGTT                                 | 1879 |
| Db | 734  | spLysSerAspGlnAlaAspAlaIleLeuArgGluAspMetGluInLeuThrGluSerS  | 754  |
| QY | 1880 | CT-----A   | 1882 |
| Db | 754  | erAsnTyrGluTrpPheAspSerGluValSerMetValLysProAspMetGlnThrLeuT | 774  |
| QY | 1883 | TGACCATTTGCCCAAGGGCCGAGG-----                                | 1908 |
| Db | 774  | rpThrIleLeuProGluSerGluArgValTrpThrArgThrSerSerLeuGluLysLeuS | 794  |
| QY | 1908 | -----  | 1908 |
| Db | 794  | erArgAspIleLeuAlaSerThrProGlnSerAlaAspArgLeuTrpLeuSerValThrG | 814  |
| QY | 1908 | -----  | 1908 |
| Db | 814  | lnSerThrLysLeuProProThrThrIleSerThrLeuLeuGluAspGluValIleMetG | 834  |
| QY | 1908 | -----  | 1908 |
| Db | 834  | lyValGlnAspIleSerLeuLeuLeuAspArgIleGlyThrAspTyrTyrGlnProGluG | 854  |
| QY | 1908 | -----  | 1908 |
| Db | 854  | lnValGlnGluGlnAsnGlyLysValGlySerTyrValGluMetSerThrSerValHisS | 874  |
| QY | 1908 | -----  | 1908 |
| Db | 874  | erThrGluMetValSerValAlaItrpProThrGluGlyGlyAspLeuSerTyrThrG   | 894  |
| QY | 1909 | -----CTGGTAGTGTCTTCAGTCTCGCTGCTACATGGCGCTCTCCCA              | 1954 |
| Db | 894  | lnThrSerGlyAlaLeuValValPhePheSerLeuArgValThrAsnMetMetPheSerG | 914  |
| QY | 1955 | ACGACCTGTTCAACAGAGCTCTCTGAGTACCGAGCTCTGGAGCAACAATTCACACAGC   | 2014 |
| Db | 914  | luAspLeuPheAsnLysAsnSerLeuGluTyrLysAlaLeuGluGlnArgPheLeuGluL | 934  |
| QY | 2015 | TCCTGGTTCATATCTACGATCCAACTTACAGGATTTTAAGCAACTTGAATTACTTAAC   | 2074 |
| Db | 934  | euLeuValProTyrLeuGlnSerAsnLeuThrGlyPheGlnAsnLeuGluIleLeuAsn  | 954  |
| QY | 2075 | TCAGAAACGGAGCTGATCTGATAGAACAAATGAGTTTGTAAGTCTGTGCGGTATA      | 2134 |
| Db | 954  | reArgAsnGlySerIleValValAsnSerArgMetLysPheAlaAsnSerValProProA | 974  |
| QY | 2135 | ACCTCACCAAGGCTGTGACCGGGCTCTGAGGATTTTCGTTCTGTGAGCCCAACAAC     | 2194 |
| Db | 974  | snValAsnAsnAlaValTyrMetIleLeuGluAspPheCysThrThrAlaTyrAsnThrM | 994  |
| QY | 2195 | TCCATCTGGAAATAGACAGCTACTCTCTCAACATGAAACGAGCTGATCAAGCAGATCCCT | 2254 |
| Db | 994  | etAsnLeuAlaIleAspLysTyrSerLeuAspValGluSerGlyAspGluAlaAsnProC | 1014 |
| QY | 2255 | GCAAGTTCTCTGCTCGGGGAATTTGCCCAATGTGTAAAGACGACCGACTGAGGAG      | 2314 |
| Db | 1014 | ysLysPheGlnAlaCysAsnGluPheSerGluCysLeuValAsnProTrpSerGlyGluA | 1034 |
| QY | 2315 | CGGAGTGTCTGTCACCAACACGAGGATATGACAGCCAG-----GGGAGCTCTG        | 2359 |
| Db | 1034 | lalsCysArgCysPheProGlyTyrLeuSerValGluGluArgProCysGlnSerLeuc  | 1054 |
| QY | 2360 | ACGCTGTGGAAACAGGCTCTGTGGCCCTGGCAACAAAGGAATCGAGGTCCTCCAGGAA   | 2419 |
| Db | 1054 | ysAspLeuGlnProAspPheCysLeuAsnAspGlyLys---CysAspIleMetProGlyH | 1073 |
| QY | 2420 | AGGAGCTCCATGCAGGTTGCCAGATC-----ACTCTGAATCAAGCATACA           | 2467 |
| Db | 1073 | isGlyAlaIleCysArg-CysArgValGlyAsnTrpTyrTrpArgGlyLysHisCys    | 1092 |

```

QY 2468 AAAGTGGTGTAAAAAGTTCACAAATCAACAAATCAACAGGTAAATCAGTAAGAGAAATT 2527
Db   :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 1093 GluGluPheValSerGluProValIleIleIleIleIleIleIleIleIleIleIleIle 1112
Db   :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 2528 CTGAATTACTGACCGTAGAATATGAGAAATTTAACCATCAAGATTGGAGGAAATTTAA 2587
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1113 Leu----- 1113

QY 2588 AACTGAAAATGTCAAAATTTACTTAGGCTATCTCAAGAGAGATGATTGGCTCTCAAG 2647
Db   :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 1114 -----ValIlePheSerAlaIleIleTyPhePheIleArgThrLeuGln 1128
Db   :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 2648 GAAATGAGACAGGCAAT-----ATCATGGGTCAATCAAAATCCAGACAT 2692
Db   :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 1129 AlaHiHisAspArgSerGluArgGluSerProPheSerGlySerSerArgGlnProAsp 1148
Db   :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 2693 ACAGTCAACACT---CAGAAAT-----CAGCACACACCATATTTCAAATATAGAGA 2740
Db   :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 1149 SerLeuSerSerIleGluAsnAlaValIleTyAsnProValTyGluSerHisArg 1167
Db   :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

RESULT 8
US-10-007-270-15
; Sequence 15, Application US/10007270
; Publication No. US20020160954A1
; GENERAL INFORMATION:
; APPLICANT: Hageman, Gregory S.
; APPLICANT: Kuehn, Markus H.
; APPLICANT: University of Iowa Research Foundation
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS FOR OCULAR ABNORMALITIES
; FILE REFERENCE: 020618-00012005
; CURRENT APPLICATION NUMBER: US/10/007,270
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 09/430,195
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 09/183,972
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 185
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Monkey species
; FEATURE:
; OTHER INFORMATION: Monkey IPM 150 amino acid sequence (partial).
US-10-007-270-15

Alignment Scores:
Pred. No.: 4.33e-61 Length: 185
Score: 841.00 Matches: 166
Percent Similarity: 85.44% Conservative: 10
Best Local Similarity: 80.58% Mismatches: 8
Query Match: 14.31% Indels: 22
DB: 13 Gaps: 1

US-10-007-270-1 (1-3330) x US-10-007-270-15 (1-185)
QY 347 TTTTCCCAACGGGGTAAAGTCTGTCCACAGGAATCCATGAAACAGATTTTACAGAGT 406
Db 2 PhePheProAsnGlyValIleValCysProGlnGluSerMetIleGlnIleLeuAlaSer 21
QY 407 CTTCAAGCTTTATATAGATTGAGAGTGTGTCCAGAGCAGATGATGGGAGCATATCGGATC 466
Db 22 LeuGlnAlaIleTyArgIleuArgValCysGlnGluAlaValTrpGluAlaIleArgIle 41
QY 467 TTTCTGGATCGCATCCCTGCACACAGGGGAATATCAGGACTGGGTCTAGCATCTGCCAGAG 526
Db 42 PheLeuAspArgIleProAspThrGlyGluTyGlnAspTrpValSerPheCysGlnGln 61
QY 527 GAGACCTTTCGCTCTTTTGACATTTGGAAAAAAATTCACGAATTCACAGAGCAGCTCGAT 586
Db 44 GAGACCTTTCGCTCTTTTGACATTTGGAAAAAAATTCACGAATTCACAGAGCAGCTCGAT 586

```

476 CGCATTCCTGCACACAGGGGAATATCAGACATGGGTGAGCATCTGCACAGCAGGACACTTC 535  
Db  
27 ArgLeuProGlyArgAspGluTyrArgHisTrpMetAsnLeuCysGluAspGlyValThr 46  
QY  
536 TGCCCTCTTTCAATTCGAAATAATTCAGCAATTTCCAGAGCACCCTGGATCTTCTCCAG 595  
Db  
47 SerValPheGluMetGlyAlaHisPheSerGlnSerValGluHisArgAsnLeuIleMet 66  
QY  
596 CAGAGAATA-----AAACAGAGAAGTTTCCTGCACGAAAGAAGTCAAAATATCTGCA 646  
Db  
67 LysLysLeuAlaTyrThrArgGluAlaGluSerSerSerCysLysAspGln---SerCys 85  
QY  
647 GAGAAGACATTCGGAGACCT-----GGTGAACCATTTGTCATTTTCAACAGCAATC 697  
Db  
86 GlyProGluLeuSerPheProValProIleGlyGluThrSerThrLeuThrGlyAlaVal 105  
QY  
698 TACATTTCAAAGACTTGGCAGCTATTTCTAGAAACCTTCAGAAGACAAATTCAGATG 757  
Db  
105 ----- 105  
QY  
758 TTGCCAACGTCCTCACTTGGGCGCTTTCCTC-----CTCACTCTCTGATGACACC 804  
Db  
106 SerSerAlaSerTyrProGlyLeuAlaSerGluSerSerAlaAlaSerProGlnGluSer 125  
QY  
805 CTCTCTCAATGAATTTCCGATTAATACACTCAACGACCAAGATGCTACAAACAGAAAGA 864  
Db  
126 IleSerAsnGluIle---GluAsnValThrGluGluProThrGlnProAlaAlaGluGln 144  
QY  
865 GAAACAGAAATTCGCTGTGTTCGACGACAGGGGTGGAGCTCAGCGTCTCTCTGGTAAAC 924  
Db  
145 IleAlaGluPhe-----SerIleGlnLeuLeuGly 154  
QY  
925 CAGAAGTTCAAGGCAGAGCTCGCTGACTCCAGTCCCATATTCACGAGCTAGCAGAGA 984  
Db  
155 LysArgTyrSerGluGluLeuArgAspProSerSerAlaLeuTyrArgLeuValGlu 174  
QY  
985 AAGTCCCACTTCAGATCGAAAGAATATTAAAGAACTTCCAGGATTCAAAAATPCCAT 1044  
Db  
175 GluPheIleSerGluValGluLysAlaPheThrGlyLeuProGlyTyrLysGlyIleArg 194  
QY  
1045 GTGTTAGCATTTAGACCAAGAAGAAAGATGGCTCAAGCTCCACAGAGATGCAACT 1104  
Db  
195 ValLeuGluPheArgAlaProGluGluAsnAspSerGlyIleAspValHisTyrAlaVal 214  
QY  
1105 ACGGCATCTTTAAGAGACACAGTCGACGAAGCAAAAGCCCTGCAGCT---GACCTCTG 1161  
Db  
215 Thr-----PheAsnGlyGluAlaIleSerAsnThrThrTrpAspLeuIle 229  
QY  
1162 TCTTTTGATTCACAAATTTGAAAGTGAAGGATCTATCATGGAACCATGACGAGAGAC 1221  
Db  
230 SerLeuHisSerAsnLysValGluAsn-----HisGlyLeuValGluMetAsp 245  
QY  
1222 AAGCAACAGAAATCTATCTCACAGCTACAGACCTCAAAAGCGCTGATCAGCAAGACCTA 1281  
Db  
246 AspLysProThrAlaValTyrThrIleSerAsnPheArgAspTyrIleAlaGluThrLeu 265  
QY  
1282 GAGGAGAGA----- 1290  
Db  
266 HisGlnAsnPheLeuMetGlyAsnSerSerLeuAsnProAspProLysProLeuGlnLeu 285  
QY  
1291 -----CAATCTTTGGATGTG----- 1305  
Db  
286 IleAsnValArgGlyValLeuLeuLeuProGlnThrGluAspIleValTrpAsnThrGlnSer 305  
QY  
1306 GGCACAATTCAGTTCACTGATGAATTCCTGGATCACTCCACGCGCTTTGGTCTTGAC--- 1362  
Db  
306 SerSerLeuGlnValThr-----ThrSerSerIle\*\*\*ValLeuGlnProAspLeu 322  
QY  
1363 ---ACCAATTCAGCTGCCCACTCTTTTGGTGTGTATTAACAGAGAGACTCTACTTTCAGT 1419  
Db  
323 ProValAlaProGluGlyArgThrSerGlySerPheIleLeuGluAspGlyLeuAlaSer 342

|    |      |   |      |
|----|------|---|------|
| QY | 1420 | CCAGAA-----   | 1425 |
| Db | 343  | ThrGluGluLeuGluAspThrSerIleAspGlyLeuProSerSerProLeuIleGlnPro  | 362  |
| QY | 1426 | -----CTTCTCTCTCTGAA-----                                      | 1443 |
| Db | 363  | ValProLysGluThrValProMetGluAspSerAspThrAlaLeuLeuSerThrPro     | 382  |
| QY | 1444 | CAGCTTGGAG-----ACAGTGGAC--GGACGACGATGTCCTA                    | 1479 |
| Db | 383  | HisLeuThrSerSerAlaIleGluAspLeuThrLysAspIleGlyThrProSerGlyLeu  | 402  |
| QY | 1480 | -----CCTGACACT  | 1488 |
| Db | 403  | GluSerLeuAlaSerAsnIleSerAspGlnLeuGluValIleProTrpPheProAspThr  | 422  |
| QY | 1489 | TCT-----  | 1491 |
| Db | 423  | SerValGluLysAspPheIlePheGluSerGlyLeuGlySerGlyLysAspVal        | 442  |
| QY | 1492 | -----TGCTCTCCACCTGCTATGCGCTCTACC-----                         | 1518 |
| Db | 443  | AspValIleAspTrpProTrpSerGluThrSerIleuGluLysThrLysProLeuSer    | 462  |
| QY | 1519 | -----TCCCTGTCAGAGCTCCA-----                                   | 1536 |
| Db | 463  | LysSerTrpSerGluGluGlnAspAlaLeuLeuProThrGluGlyArgGluLysLeuHis  | 482  |
| QY | 1536 | -----   | 1536 |
| Db | 483  | IleAspGlyArgValAspSerThrGluGlnIleIleGluSerSerGluHisArgTyrGly  | 502  |
| QY | 1537 | -----CCTTTCTTTT   | 1545 |
| Db | 503  | AspArgProIleHisPheIleGluGlu***SerHisValArgSerThrIleProIlePhe  | 522  |
| QY | 1546 | ATGGCATCAAGC-----ATCTCTCT-----CTGACTGATCAAGGC                 | 1581 |
| Db | 523  | ValGluSerAlaThrProProThrSerProIlePheSerLysHisThrSerAspValPro  | 542  |
| QY | 1582 | ACCACAGATACAATGGCCACTGACACGACATGCTAGTACCA---GGGCTCACATPCCCC   | 1638 |
| Db | 543  | AspIleAspSerTyrSerLeuThrLysProPheLeuProValThrIleAlaIlePro     | 562  |
| QY | 1639 | ACCAGTGAATTATCTGCAATCAGCAACCTGCTCGGAAATTCACATCCACCTGCATCT     | 1698 |
| Db | 563  | AlaSerThrLysLysThrAspGluValLeuLysGluAspMetValHisThrGluSerSer  | 582  |
| QY | 1699 | TCAGAT-----GACAGCGCATCAAGTCGAGTGGCGAGATATG-----               | 1737 |
| Db | 583  | SerHisLysGluLeuAspSerGluValProValSerArgProAspMetGlnProValTrp  | 602  |
| QY | 1738 | -----GTCAGACACCTAGAT  | 1752 |
| Db | 603  | ThrMetLeuProGluSerAspThrValTrpThrArgThrSerSerLeuGlyLysLeuSer  | 622  |
| QY | 1753 | GAAATGGATCTGCTGACACTCTCT-----                                 | 1776 |
| Db | 623  | ArgAspThrLeuAlaSerThrProGluSerThrAspArgLeuTrpIleLysAlaSerMet  | 642  |
| QY | 1777 | ---GCCCATCTGAGGTACCA-----GAGCTCAGCGAATATGTTTCT                | 1815 |
| Db | 643  | ThrGlnSerThrGluLeuProSerThrThrHisSerThrGlnLeuGluGluValIle     | 662  |
| QY | 1816 | GTC-----  | 1818 |
| Db | 663  | MetAlaValGlnAspIleSerIleuGluLeuAspGlnValGlyThrAspTyrThrGlnSer | 682  |
| QY | 1819 | -----CCAGATCATTTCTTGAGGATACACCTCTGTC                          | 1851 |
| Db | 683  | GluLeuThrGluGluGlnHisGlyLysAlaAspSerTyrValGluMetSerThrSerVal  | 702  |
| QY | 1852 | TCAGCTTTACAGTATATACCACTAGTGTCTATGACCATTTGCCCCCAAGGCG-----     | 1902 |





```

; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersch, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 158
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-158

Alignment Scores:
Pred. No.: 1-62e-06 Length: 1255
Score: 189.50 Matches: 92
Percent Similarity: 41.48% Conservative: 54
Best Local Similarity: 26.14% Mismatches: 131
Query Match: 3.23% Indels: 75
DB: 14 Gaps: 19

US-10-007-270-1 (1-3330) x US-10-171-311-158 (1-1255)
QY 1345 CCAGCTTTGGT-----CCTGACCAATCAGAGTCCACATCTTTGCT 1392
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
814 ProAlaHisGlyValThrSerAlaProAspThrArg---ProAlaProGlySerThrAla 832
QY 1393 GTTATAACAGAGGATGCTACTTTGAGTCCAGACTCTCTCTGTTGACCCAGCTTGAG 1452
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
833 ProProAlaHisGlyValThrSerAlaProAspThrArgProAlaProGly-----Ser 850
QY 1453 ACAGTGGAGGAGCAGAGCATGCTCTA-----CCTGACACT----- 1498
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
851 ThrAlaProProAla---HisGlyValThrSerAlaProAspThrArgProAlaProGly 869
QY 1489 TCTTGCTCTCCACTGCTATGCGCTCTACTCCCTG----- 1524
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
870 SerThrAlaProProAlaHisGlyValThrSerAlaProAspThrArgProAlaProGly 889
QY 1525 TCAGAGCTCCACTCTTCTTTATGGCATCAGCATCTTCTCTGACTGAT----- 1575
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
890 SerThrAlaProPro-----AlaHisGlyValThrSerAlaProAspThrArgPro 906
QY 1576 ---CAAGGACACACAGAT-----ACAATGGCCACTGACACAGCAATG 1614
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
907 AlaProGlySerThrAlaProProAlaHisGlyValThrSerAlaProAsp---ThrArg 925
QY 1615 CTAGTACAGGCTCCACTCCACCCACAGT-----GATTATCTGCAATCAAGCAA 1665
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
926 ProAlaProGlySerThrAlaProProAlaHisGlyValThrSerAlaProAspAsnArg 945
QY 1666 CTGGCTCTGGGAATTTACATCCACT-----GCATCTTCAGATGACGCGCATCA 1716
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
946 ProAlaLeuGlySerThrAlaProProValHisAsnValThrSerAlaSerGlySerAla 965
QY 1717 AGTGCAGTGGCAGATATGCTCAGACACTAGATGAATGGATCTGCTGACACTCT 1776
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
966 SerGlySerAlaSerThrLeuValHisAsnGlyThrSerAlaArgAlaThrThrPro 985
QY 1777 GCCCATCTGAGGTACAGAGCTCAGCAATATGTTCTGTCACGATCTTCTTGAG 1836
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
986 AlaSerGlySerThrPro-----PheSerIleProSerHisSerAsp 1000
QY 1837 GATACCACTCTCTGTC-----TCA 1854

```

```

Db 1001 ThrProThrThrLeuAlaSerHisSerThrLysThrAspAlaSerSerThrHisHisSer 1020
QY 1855 GCTTACAGTATATACCACTAGTTCTATGACACTTGGCCCAAGGGCCGAGAGCTGGTA 1914
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1021 SerValProProLeuThrSerSerAsnHisSerThrSerProGlnLeuSerThrGlyVal 1040
QY 1915 GTGTTCTTC-----AGTCTGCTGCTGCTTAACATGGCTTCTTCCACGACCTGTTCAAC 1968
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1041 SerPhePhePheLeuSerPheHisIleSerAsnLeuGlnPheAsnSerSerLeuGluAsp 1060
QY 1969 AAGAGCTCTCGAGTACCGAGCTCTGAGGACCAATTCACACAGCTGCTGCTTCATAT 2028
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1061 ProSerThrAspTyrTyrGlnGlnLeuGlnArgAspIleSerGluMetPheLeuGlnIle 1080
QY 2029 CTACAGTCAATCTTACAGGATTTAAGCACTTGAATTAATCTTAACTTCAGAACGGAGT 2088
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1081 TyrLysGlnGly-----GlyPheLeuGlyLeuSerAsnIleLysPheArgProGlySer 1098
QY 2089 GTGATTGTGAATAGCAAAATGAAGTTTGTAAGTCTGTGCGGTATTAACCTCACCAGGCT 2148
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1099 ValValGlnLeuThrLeuAlaPheArgGluGly-----ThrIleAsn 1113
QY 2149 GTGACGCGGCTC-----TTGAGGATTTTTCGTTCTGCTGCGAGCCCAACACTCCAT 2199
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1114 ValHisAspValGluThrGlnPheAsnGlnTyrLysThrGluAlaAlaSerArgTyrAsn 1133
QY 2200 CTGGAATATAGACACTACTCTCTCAACATTGAACCA 2235
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1134 LeuThrIleSerAspValSerValSerAspValPro 1145

```

## RESULT 13

```

US-10-177-293-311
; Sequence 311, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Puztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0

```



```

; SEQ ID NO 311
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-311

Alignment Scores:
Pred. No.: 1,62e-06 Length: 1255
Score: 189.50 Matches: 92
Percent Similarity: 41.48% Conservative: 54
Best Local Similarity: 26.14% Mismatches: 131
Query Match: 3.23% Indels: 75
DB: 14 Gaps: 19

US-10-007-270-1 (1-3330) x US-10-177-293-311 (1-1255)

QY 1345 CAGGCTTTGGT-----CCTGACACCAATCAGAGCTGCCACATCTTTTGGT 1392
Db ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
814 ProAlaHisGlyValThrSerAlaProAspThrArg---ProAlaProGlySerThrAla 832
QY 1393 GTTATAACAGAGGAGTCTACTTTGAGTCCAGAACTCCCTCTGTTGAACCCACGCTTGAG 1452
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
833 ProProAlaHisGlyValThrSerAlaProAspThrArgProAlaProGly-----Ser 850
QY 1453 ACAGTGGACGAGCAGACATGGTCTA-----CCTGACACT-----1488
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
851 ThrAlaProProAla--HisGlyValThrSerAlaProAspThrArgProAlaProGly 869
QY 1489 TCTTGGTCTCCACTGCTGATGGCTTACTCTCCCTG-----1524
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
870 SerThrAlaProProAlaHisGlyValThrSerAlaProAspThrArgProAlaProGly 889
QY 1525 TCAGAAGCTCCACCTTCTTTATGGCATCAAGCACTTCTCTGACTCAT-----1575
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
890 SerThrAlaProPro-----AlaHisGlyValThrSerAlaProAspThrArgPro 906
QY 1576 ----CAAGCCACACAGAT-----ACAATGGCCACTCACCAGCAATG 1614
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
907 AlaProGlySerThrAlaProProAlaHisGlyValThrSerAlaProAsp---ThrArg 925
QY 1615 CTAGTACAGGGCTCAGCATCCACCCACAGT-----GATTATCTGCAATCAGCAA 1665
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
926 ProAlaProGlySerThrAlaProProAlaHisGlyValThrSerAlaProAspAsnArg 945
QY 1666 CTGCTCTGGGAATTCATCCACT-----GCATCTCAGATCAGACCGCATCA 1716
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
946 ProAlaLeuGlySerThrAlaProProValHisAsnValThrSerAlaSerGlySerAla 965
QY 1717 AGTCAGGTGGCGAAGATATGTCGACAGACACCTAGATGAATGGATCTCTGACACTCTCT 1776
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
966 SerGlySerAlaSerThrLeuValHisAsnGlyThrSerAlaArgAlaThrThrPro 985
QY 1777 GCCCATCTGAGTACAGAGCTCAGCAATATGTTCTGCCAGATCAATTTCTGGAG 1836
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
986 AlaSerGlySerThrPro-----PheSerIleProSerHisHisSerAsp 1000
QY 1837 GATACCACTCTCTGTC-----TCA 1854
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1001 ThrProThrThrLeuAlaSerHisSerThrLysThrAspAlaSerSerThrHisHisSer 1020
QY 1855 GCTTTACATATATACACACTAGTCTTATGACCAATTCGCCCAAGGCGCAGAGCTGGTA 1914
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1021 SerValProProLeuThrSerSerAsnHisSerThrSerProGlnLeuSerThrGlyVal 1040
QY 1915 GTGTTCTTC-----AGTCGCGTGTCTAAGATGGCTTCTCCAAAGCCTGTTCAAC 1968
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1041 SerPhePheLeuSerPheHisIleSerAsnLeuGlnPheAsnSerSerLeuGluAsp 1060
QY 1969 AAGAGCTCTCTGGAGTACCGAGCTCTGAGCAACAATTCACACAGCTGCTGTTCCATAT 2028
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1061 ProSerThrAspTyrTyrGlnLeuGlnArgAspIleSerGluMetPheLeuGlnIle 1080
QY 2029 CTACGATCGAATCTTACAGGATTTAAGCAACTTGAATACTTAATCTCAGAAAGGGAGT 2088

```

```

Db 1081 TyrLysGlnGly-----GlyPheLeuGlyLeuSerAsnIleLysPheArgProGlySer 1098
QY 2089 GTGATTCTGAATAGCAAAATGAAGTTTGTCTAGTCTTGCGGTATTAACCTCACAAGGCT 2148
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1099 ValValValGlnLeuThrLeuAlaPheArgGlnGly-----ThrIleAsn 1113
QY 2149 GTGCACGGGGTC-----TTGGAGGATTTTCTGCTGCGAGCCCAACAACCTCCAT 2199
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1114 ValHisAspValGluThrGlnPheAsnGlnTyrLysThrGluAlaAlaSerArgTyrAsn 1133
QY 2200 CTGGAATAGACAGCTACTCTCTCAACATGACCA 2235
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1134 LeuThrIleSerAspValSerValSerAspValPro 1145

RESULT 14
US-10-097-340-212
; Sequence 212, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPARU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAWATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISSEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; FILE REFERENCE: Assessment, Prevention, and Therapy of Ovarian Cancer
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 212
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-340-212

Alignment Scores:
Pred. No.: 1.33e-06 Length: 515
Score: 188.50 Matches: 85
Percent Similarity: 40.73% Conservative: 60
Best Local Similarity: 23.98% Mismatches: 146
Query Match: 3.21% Indels: 65
DB: 14 Gaps: 16

US-10-007-270-1 (1-3330) x US-10-097-340-212 (1-515)

```

```

QY 1291 CAATCTTTGGATGTTGGGACAAATTCAGTTCATGATGAATTCCTGGATCACTCCAGCC 1350
Db |||||
74 GlnGlyGlnAspValThrLeuAlaProAlaThrGluProAlaSerGlySerAlaAlaThr 93
QY 1351 TTTGGTCTCGACACCAATCAGAGCTGCCACATCTCTTTGCTGTATA----- 1398
Db |||||
94 TrpGlyGlnAspValThrSer---ValProValThrArgProAlaLeuGlySerThrThr 112
QY 1399 -----ACAGAGATGCTACTTTGAGTCCAGAACTCTCTCTGTGAACCCAGCTTGAG 1452
Db |||||
113 ProProAlaHisAspValThrSerAlaProAspAsnLysProAlaProGly-----Ser 130
QY 1453 ACAGTGGAGCGAGCAGCATGGTCTA-----CCTGACACT----- 1488
Db |||||
131 ThrAlaProProAla---HisGlyValThrSerAlaProAspThrArgProProGly 149
QY 1489 TCTTGCTCTCCACCTGCTATGGCTCTACCTCCCTGTCAGAACT-----CCACCTTTC 1542
Db |||||
150 SerThrAlaProAlaHisGlyValThrSerAlaProAspThrArgProProGly 169
QY 1543 TTTATGGCATCAACATCTCTCTGACTGATCAAGGCCACACAGATACATGGCCACT 1602
Db |||||
170 SerThrAlaProAlaAlaHisGlyValThr-----SerAla 181
QY 1603 GACCGACAAATGCTAGTACCGGCTCACCATCCCAACAGCT-----GATTATCT 1653
Db |||||
182 ProAspThrArgProAlaProGlySerThrAlaProProAlaHisGlyValThrSerAla 201
QY 1654 GCAATCAGCCAACTGGCTCTGGAAATTCATCCACT-----GCATCTTCAGAT 1704
Db |||||
202 ProAspAsnArgProAlaLeuAlaSerThrAlaProProValHisAsnValThrSerAla 221
QY 1705 GACAGCCGATCAAGTGCAGGTGGGAGATATGTCACAGACATAGATGAATGATCTG 1764
Db |||||
222 SerGlySerAlaSerGlySerAlaSerThrLeuValHisAsnGlyThrSerAlaArgAla 241
QY 1765 TCTGACACTCTGCCCCAATCTGAGTACAGAGCTCAGCGAATATCTTTGCTCCAGAT 1824
Db |||||
242 ThrThrThrProAlaSerLysSerThrPro-----PheSerIleProSer 256
QY 1825 CATTTCTTGAGGATACCACTCTCTC----- 1851
Db |||||
257 HisHisSerAspProThrThrLeuAlaSerHisSerThrLysThrAspAlaSerSer 276
QY 1852 -----TCAGCTTTACAGTATACCACTAGTCTATGACATGTCGCCCAAGGGC 1902
Db |||||
277 ThrHisSerThrValProProLeuThrSerSerAsnHisSerThrSerProGlnLeu 296
QY 1903 CGAGAGCTGTAGTCTCTC-----AGTCTGCGTGTGCTAAATGCGCTCTCCAAAC 1956
Db |||||
297 SerThrGlyValSerPhePheLeuSerPheHisLeuSerAsnLeuGlnPheAsnSer 316
QY 1957 GACCTGTTCAACAGAGCTCTCTGGAGTACGAGCTCTGGAGCAACAAATTCACAGCTG 2016
Db |||||
317 SerLeuGluAspProSerThrAspTyrTyrGlnGlnLeuGlnArgAspIleSerGluMet 336
QY 2017 CTGGTTCATATCTACGATCAATCTTACAGATTAAAGCAACTGGAATCACTTACTTC 2076
Db |||||
337 PheLeuGlnIleTyrLysGlnGly-----GlyPheLeuGlyLeuSerAsnIleLysPhe 354
QY 2077 AGAAGCGAGTGTGATGTGAATAGCAAAATGAAGTTTGCTAAGTCTGTCGCGTATAC 2136
Db |||||
355 ArgProGlySerValValGlnLeuThrLeuAlaPheArgGluGly----- 370
QY 2137 CTCACCAAGGCTGTCACGGGTC-----TTGAGGATTTTCGTTCTGTCGAGCC 2187
Db |||||
371 ---ThrIleAsnValHisAspValGluThrGlnPheAsnGlnTyrLysThrGluAlaAla 389
QY 2188 CAACAACTCCATCTGGAATAGACAGCTACTCTCTCAACTTGAACCA 2235
Db |||||
390 SerArgTyrAsnLeuThrIleSerAspValSerValSerAspValPro 405

```

RESULT 15

```

US-10-171-311-156
; Sequence 156, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Ganavarapu, Manjula
; APPLICANT: Hoerish, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 156
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-156

```

## Alignment Scores:

```

Pred. No.: 1.33e-06 Length: 515
Score: 188.50 Matches: 85
Percent Similarity: 40.73% Conservative: 60
Best Local Similarity: 23.88% Mismatches: 146
Query Match: 3.21% Indels: 65
DB: 14 Gaps: 16

```

US-10-007-270-1 (1-3330) x US-10-171-311-156 (1-515)

```

QY 1291 CAATCTTTGGATGTTGGGACAAATTCAGTTCATGATGAATTCCTGGATCACTCCAGCC 1350
Db |||||
74 GlnGlyGlnAspValThrLeuAlaProAlaThrGluProAlaSerGlySerAlaAlaThr 93
QY 1351 TTTGGTCTCGACACCAATCAGAGCTGCCACATCTCTTTGCTGTATA----- 1398
Db |||||
94 TrpGlyGlnAspValThrSer---ValProValThrArgProAlaLeuGlySerThrThr 112
QY 1399 -----ACAGAGATGCTACTTTGAGTCCAGAACTCTCTCTGTGAACCCAGCTTGAG 1452
Db |||||
113 ProProAlaHisAspValThrSerAlaProAspAsnLysProAlaProGly-----Ser 130
QY 1453 ACAGTGGAGCGAGCAGCATGGTCTA-----CCTGACACT----- 1488
Db |||||
131 ThrAlaProProAla---HisGlyValThrSerAlaProAspThrArgProProGly 149
QY 1489 TCTTGCTCTCCACCTGCTATGGCTCTACCTCCCTGTCAGAACT-----CCACCTTTC 1542
Db |||||
150 SerThrAlaProAlaHisGlyValThrSerAlaProAspThrArgProProGly 169
QY 1543 TTTATGGCATCAACATCTCTCTGACTGATCAAGGCCACACAGATACATGGCCACT 1602
Db |||||
170 SerThrAlaProAlaAlaHisGlyValThr-----SerAla 181
QY 1603 GACCGACAAATGCTAGTACCGGCTCACCATCCCAACAGCT-----GATTATCT 1653
Db |||||
182 ProAspThrArgProAlaProGlySerThrAlaProProAlaHisGlyValThrSerAla 201
QY 1654 GCAATCAGCCAACTGGCTCTGGAAATTCATCCACT-----GCATCTTCAGAT 1704
Db |||||
202 ProAspAsnArgProAlaLeuAlaSerThrAlaProProValHisAsnValThrSerAla 221

```

Search completed: March 4, 2004, 18:59:38  
Job time : 160 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 4, 2004, 18:39:06 ; Search time 54.5 seconds  
(without alignments)  
11754.781 Million cell updates/sec

Title: US-10-007-270-1  
Perfect score: 5875  
Sequence: 1 taacacgaaggtatct.....tactatatgacataatcaat 3330

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.spool\_p/US10007270/runat\_04032004\_160717\_5031/app\_query.fasta\_1.3527  
-Q=/cgn2.1/USP70.spool\_p/US10007270/runat\_04032004\_160717\_5031/app\_query.fasta\_1.3527  
-DB=PIR\_78 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -ALIGN=15 -MODE=LOCAL  
-USER=US10007270@cgn 1.1.84 @runat\_04032004\_160717\_5031 -NCPV=6 -ICPU=3  
-NO MAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR 78.\*

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| Result No. | Score | Query Match | Length | DB ID    | Description          |
|------------|-------|-------------|--------|----------|----------------------|
|            |       |             |        |          |                      |
| 1          | 188.5 | 3.2         | 1344   | 1 A35175 | mucin 1 precursor    |
| 2          | 160.5 | 2.7         | 3507   | 2 T34513 | hypothetical prote   |
| 3          | 157   | 2.7         | 1621   | 2 A82255 | hypothetical prote   |
| 4          | 155.5 | 2.6         | 768    | 2 H81236 | DNA topoisomerase    |
| 5          | 153   | 2.6         | 919    | 2 C86578 | type C/Gen secretion |
| 6          | 153   | 2.6         | 919    | 2 F72045 | type III secretion   |
| 7          | 153   | 2.6         | 3678   | 2 S28916 | dystrophin - mouse   |
| 8          | 152.5 | 2.6         | 768    | 2 A82009 | DNA topoisomerase    |
| 9          | 152.5 | 2.6         | 2397   | 1 A55535 | versican precursor   |
| 10         | 150.5 | 2.6         | 595    | 2 G64529 | outer membrane pro   |
| 11         | 149.5 | 2.5         | 631    | 2 I52517 | episialin - mouse    |
| 12         | 148.5 | 2.5         | 1634   | 2 T26517 | hypothetical prote   |
| 13         | 148   | 2.5         | 589    | 2 S74668 | hypothetical prote   |
| 14         | 147   | 2.5         | 1367   | 1 S48478 | glucan 1,4-alpha-g   |

#### SUMMARIES

15 147 2.5 1620 2 S61535 nucleotide-binding  
16 147 2.5 1751 2 A45604 major blood-stage  
17 146.5 2.5 2409 1 A60979 versican precursor  
18 146 2.5 1280 2 T00365 hypothetical prote  
19 146 2.5 1285 2 H85041 hypothetical prote  
20 145.5 2.5 809 2 T39626 hypothetical prote  
21 145 2.5 630 2 A39344 tumor-associated m  
22 145 2.5 1036 2 S73601 protein P200 - Myc  
23 145 2.5 1864 2 F86378 protein F2109.12 [  
24 145 2.5 1957 2 T38077 hypothetical coile  
25 145 2.5 2484 2 T26216 hypothetical prote  
26 145 2.5 2607 2 T26216 hypothetical prote  
27 145 2.5 2829 2 A42771 reticulocyte-bind  
28 144 2.5 1468 2 A4345 nucleoporin - rat  
29 143.5 2.4 1306 2 S25370 MS2 protein - yea  
30 143 2.4 1120 2 JCT765 mitotic spindle as  
31 143 2.4 3225 2 I52300 giantin - human  
32 143 2.4 3259 1 A56539 giantin - human  
33 142.5 2.4 347 2 S10571 mucin 1 precursor,  
34 142.5 2.4 772 2 T27907 hypothetical prote  
35 142.5 2.4 1113 2 T47381 hypothetical prote  
36 141.5 2.4 3381 2 T42389 versican precursor  
37 141 2.4 2845 2 I49505 adenomatous polyp  
38 141 2.4 3685 1 A27605 dystrophin, muscle  
39 140.5 2.4 529 2 A41137 heat shock transcr  
40 140.5 2.4 927 2 T09006 scnM protein - Str  
41 140.5 2.4 1772 2 A45532 major neurozoite su  
42 139.5 2.4 1928 2 S46773 myosin heavy chain  
43 139.5 2.4 2357 2 A59249 class VII unconven  
44 139 2.4 2145 2 JCA747 adenylate cyclase  
45 138.5 2.4 3187 2 JCS937 364K Golgi complex

ALIGNMENTS

RESULT 1  
A35175  
mucin 1 precursor, repetitive splice form A [validated] - human  
N:Alternate names: breast carcinoma-associated DF3 antigen; core protein KP39; episialin.  
ncretatic mucin; polymorphic epithelial mucin (PEM)  
C:Species: mucin 1 precursor, epithelial tumor antigen splice form; mucin 1 precursor, i  
J:Biol. Chem. 265, 5573-5578, 1990  
C:Accession: A35175; B51575; A35886; A35887; S10572; S40293; A36735; PX0066; S10218; S51  
R:Litgenberg, M.J.L.; Vos, H.L.; Gemmisen, A.M.C.; Hilkens, J.  
A:Title: Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene enc  
A:Reference number: A35175; MUID:90202794; PMID:2318825  
A:Accession: A35175  
A:Molecule type: mRNA  
A:Residues: 1-952,1033-1344 <LIG1>  
A:Cross-references: GB:M32738; GB:J05288; NID:G182121; PID:AAA35804.1; PID:G182124; GB:N  
A:Experimental source: splice form A  
A:Note: GenBank entries HUMEPIS1A1 and HUMEPIS1A2 present only the amino-and carboxyl-enc  
A:Accession: B35175  
A:Molecule type: mRNA  
A:Residues: 1-19,29-952,1033-1344 <LIG2>  
A:Cross-references: GB:M32739; GB:J05288; NID:G182126; PID:AAA35806.1; PID:G182129; GB:N  
A:Experimental source: splice form B  
A:Note: GenBank entries HUMEPIS1B1 and HUMEPIS1B2 present only the amino-and carboxyl-enc  
J:Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Burchell  
J. Biol. Chem. 265, 15286-15293, 1990  
A:Title: Molecular cloning and expression of human tumor-associated polymorphic epitheli  
A:Reference number: A35886; MUID:90368715; PMID:1697589  
A:Accession: A35886  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-19,29-992,1033-1344 <GEN>  
A:Cross-references: GB:J05581; NID:G188869; PID:AAA59876.1; PID:G188870  
A:Note: GenBank entry HUMUCAB includes one copy of the tandemly repeated sequence  
R:lan, W.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.  
J. Biol. Chem. 265, 15294-15299, 1990  
A:Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.

A:Reference number: A35887; MUID:90368716; PMID:2394722  
 A:Accession: A35887  
 A>Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>  
 A:Cross-references: GB:J05582; NID:G189598; PIDN:AAA60019.1; PID:G189599  
 A>Note: GenBank entry HUMPANMU contains four fewer copies of the tandemly repeated sequence than the other entries.  
 R:Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.; Eur. J. Biochem. 189, 463-473, 1990  
 A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may generate multiple protein variants.  
 A:Reference number: S10571; MUID:90276413; PMID:2351132  
 A:Accession: S10572  
 A:Molecule type: mRNA  
 A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <WRE>  
 A:Cross-references: ENBL:X52229; NID:G37053; PIDN:CAA36478.1; PID:G37054  
 R:Abu, M.; Siddiqui, J.; Kufe, D.; Biochem. Biophys. Res. Commun. 165, 644-649, 1989  
 A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated protein.  
 A:Reference number: A36735; MUID:90088473; PMID:2597151  
 A:Accession: A36735  
 A:Molecule type: mRNA  
 A:Residues: 1-142,'Q',144-162,'Q',164-168 <ABE>  
 A:Cross-references: ENBL:M31823; NID:G181542; PIDN:AAA35757.1; PID:G181543  
 R:Masuzawa, Y.; Miyachi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H.; J. Biochem. 112, 609-615, 1992  
 A:Title: A novel core protein as well as a polymorphic epithelial mucin carry peanut agglutinin binding site.  
 A:Reference number: JX0235; MUID:93123189; PMID:1478919  
 A:Accession: JX0235  
 A:Molecule type: mRNA  
 A:Residues: 998-1011,'ES',1014-1017,1018-1032,'T',1034-1037,1038-1057 <MAS>  
 A:Experimental source: Gastric carcinoma cell  
 R:Zrihan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.; FEBS Lett. 356, 130-136, 1994  
 A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine receptor.  
 A:Reference number: S51026; MUID:95080414; PMID:7988707  
 A:Accession: S51026  
 A:Contents: annotation  
 A>Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region are partial repeats. This protein is length polymorphic. Individuals may have between 21 and 125 partial repeats. The repeat shown is defined by SmaI nuclease sites.  
 C:Comment: Serine and threonine residues in the tandem repeat domain are extensively glycosylated.  
 C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146  
 C:Genetics:  
 A:Gene: GDB:MUC1; PUM  
 A:Cross-references: GDB:120705; OMIM:158340  
 A:Map position: 1q21-q23  
 A:Introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3  
 C:Superfamily: polymorphic epithelial mucin  
 C:Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphism  
 F1-1344/Product: mucin 1 precursor, splice form A #status predicted <PRED>  
 F1-62/Region: mucin 1 amino-terminal non-repetitive  
 F1-23/Domin: signal sequence #link PRED #status predicted <SIGA>  
 F1-19,29-32/Domin: signal sequence #link PRED #status predicted <SIGB>  
 F1-19,29-1344/Product: mucin 1 precursor, splice form B #status predicted <SIGB>  
 F1-15,29-32-1033-1344/Product: mucin 1 precursor, splice form C #status predicted <PRED>  
 F1-138-1017/Region: 20-residue repeats (GSTAPPAGVTSAPDTPAP)  
 F1-143-1344/Region: mucin 1 carboxyl-terminal non-repetitive  
 F1-1245-1272/Domin: transmembrane #status predicted <TRM>  
 F1-1046,1064,1118,1144,1222/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F1-1213/Binding site: phosphate (Tyr) (covalent) #status predicted

## Alignment Scores:

|                        |          |               |      |
|------------------------|----------|---------------|------|
| Pred. No.:             | 0.000187 | Length:       | 1344 |
| Score:                 | 188.50   | Matches:      | 92   |
| Percent Similarity:    | 41.19%   | Conservative: | 53   |
| Best Local Similarity: | 26.14%   | Mismatches:   | 132  |
| Query Match:           | 3.21%    | Indels:       | 75   |

| DB:  | 1  | Gaps: | 19 |
|--|--|-------|----|
| US-10-007-270-1 (1-3330) x A35175 (1-1344) |  |       |    |
| QY   | 1345 CCAGCTTTGGT-----CCTGACACCAATCAGAGCTGCCACATCTTTTGGT 1392           |       |    |
| Db   | 903 ProAlaHisGlyValThrSerAlaProAspThrArg---ProAlaProGlySerThrAla 921   |       |    |
| QY   | 1393 GTTATACAGAGAGTACTTGTGAGTCCAGACATCTCTCTGTGTGAACCCAGCTTGAG 1452     |       |    |
| Db   | 922 ProProAlaHisGlyValThrSerAlaProAspThrArgProAlaProGly---Ser 939      |       |    |
| QY   | 1453 ACAGTGCAGCGAGCAGAGCATGGTCTA-----CCTGACACT- 1488                   |       |    |
| Db   | 940 ThrAlaProProAla---HisGlyValThrSerAlaProAspThrArgProAlaProGly 958   |       |    |
| QY   | 1489 TCTTGTCTCCACCTGCTATGGCTCTACTCCCTG----- 1524                       |       |    |
| Db   | 959 SerThrAlaProProAlaHisGlyValThrSerAlaProAspThrArgProAlaProGly 978   |       |    |
| QY   | 1525 TCAGAGCTCCACTTTCTTTTATGTCATCAAGCATCTCTCTCACTGAT----- 1575         |       |    |
| Db   | 979 SerThrAlaProPro-----AlaHisGlyValThrSerAlaProAspThrArgPro 995       |       |    |
| QY   | 1576 ---CAAGGCACACAGAT-----ACATGGCCACTGACAGCAGCAATG 1614               |       |    |
| Db   | 996 AlaProGlySerThrAlaProProAlaHisGlyValThrSerAlaProAsp---ThrArg 1014  |       |    |
| QY   | 1615 CTAGTACAGCGGTCCACATCCCAACAGT-----GATTATCTGCAATCAGCCAA 1665        |       |    |
| Db   | 1015 ProAlaProGlySerThrAlaProProAlaHisGlyValThrSerAlaProAspAsnArg 1034 |       |    |
| QY   | 1666 CTGGCTCTGGGAATTCATCCACCT-----GCATCTTCAGATCAGACCGCATCA 1716        |       |    |
| Db   | 1035 ProAlaLeuGlySerThrAlaProProValHisAsnValThrSerAlaSerGlySerAla 1054 |       |    |
| QY   | 1717 AGTCAGGTGGGAGATATGTCAGACACCTAGATGAATGATGTCGTCTGACACTCTCT 1776     |       |    |
| Db   | 1055 SerGlySerAlaSerThrLeuValHisAsnGlyThrSerAlaArgAlaThrThrPro 1074    |       |    |
| QY   | 1777 GCCCATCTGAGTACCAGCTCAGCGAATATGTTCTGCCAGATCATTTCTTGGAG 1836        |       |    |
| Db   | 1075 AlaSerGlySerThrPro-----PheSerIleProSerHisSerAsp 1089              |       |    |
| QY   | 1837 GATACACCTCTGTC----- 1854  |       |    |
| Db   | 1090 ThrProThrThrLeuAlaSerHisSerThrThrAspAlaSerSerThrHisSer 1109       |       |    |
| QY   | 1855 GCITTCAGTATATCACCACCTAGTCTTATGACCATTCGCCCAAGGGCGAGCTGTTA 1914     |       |    |
| Db   | 1110 ThrValProProLeuThrSerSerAsnHisSerThrSerProGlnLeuSerThrGlyVal 1129 |       |    |
| QY   | 1915 GTGTTCTTC-----AGTCTGCGTGTGCTAACATGACCTTCCCAAGCAGCTGTTCAAC 1968    |       |    |
| Db   | 1130 SerPhePheLeuSerPheHisIleSerAsnLeuGlnPheAsnSerSerLeuGluAsp 1149    |       |    |
| QY   | 1969 AAGAGCTCTCTGGAGTACCGAGCTCTGGAGCAACATTCACAGCTCTGGTTCATAT 2028      |       |    |
| Db   | 1150 ProSerThrAspTyrTyrGlnGlnLeuGlnArgAspIleSerGluMetPheLeuGlnIle 1169 |       |    |
| QY   | 2029 CTACGATCCAACTTACAGGATTAAGCACTTGAATATCTTAACCTCAGAAACGGAGT 2088     |       |    |
| Db   | 1170 TyrLysGlnGly-----GlyPheLeuGlyLeuSerAsnIleLysPheArgProGlySer 1187  |       |    |
| QY   | 2089 GTGATTTGATAGCAAAATGAAGTTTCTAGTCTGTGCCGTATACCTCCACCAAGGCT 2148     |       |    |
| Db   | 1188 ValValValGlnLeuThrLeuAlaPheArgGluGly-----ThrIleAsn 1202           |       |    |
| QY   | 2149 GTGACAGGGGTC-----TTGAGAGATTTTGGTCTGTGGCGCCCAACTTCAT 2199          |       |    |
| Db   | 1203 ValHisAspValGluThrGlnPheAsnGlnTyrLysThrGluAlaAlaSerArgTyrAsn 1222 |       |    |
| QY   | 2200 CTGGAAATAGACAGCTACTCTCTCAACATTGAACCA 2235                         |       |    |

```

Db      1223 LeuThrlleSerAspValSerValSerAspValPro 1234
RESULT 2
T34513
hypotheical protein ZK783.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34513
R:Pavello, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A:Description: The sequence of C. elegans cosmid ZK783.
A:Reference number: Z21536
A:Accession: T34513
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-3507 <FAV>
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021.1; CESP:ZK783.1
A:Experimental source: strain Bristol N2; clone ZK783
C:Genetics:
C:Gene: CESP:ZK783.1
A:Map position: 3
A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;
3504/1

Alignment Scores:
Pred. No.:      0.0165      Length:      3507
Score:          160.50      Matches:      128
Percent Similarity: 33.23%      Conservative: 83
Best Local Similarity: 20.16%      Mismatches: 230
Query Match:      2.73%      Indels:      194
DB:              2          Gaps:      22

US-10-007-270-1 (1-3330) x T34513 (1-3507)

QY      486 ACACAGGGGAATATCAGGATGGGTGAGCATCTGCCAGCAGAGACCTTCTGCTCTTTG 545
Db      1969 ThrlysglyileValSerSerThrAlaThrSerSer
QY      546 ACATTGGAAAACCTCAGCAATTCACGAGACCTCGGATCTTCCAGCAGAGAAATAA 605
Db      1982 -----GluSerThrThrAlaGluProHisValThrThrSerileSerSerThrThrSer 1999
QY      606 AACAGAGAGTTCCTCCACAGAGAAAGATGAATATCTGCAGAGAGACATGGGAGAGC 665
Db      2000 ThrlyaspMetThr-----Ser 2005
QY      666 CTGGTGAACCAATTCATTCACAGCAATCTACATTTCAAAGACTTGGGACATTTCT 725
Db      2006 SerlySerProGluAsnValThrMetSerSerGluSerProGluValSerThrSerSer 2025
QY      726 AAGAAACCTCAGAGAGCAAAATTCAGATGTTGCCAAGTCTCACTGGGCGCTTCCC 785
Db      2026 SerlySerThrThrAlaSerGluThr-ThrValSerSer----- 2038
QY      786 TCTACTCTGATGACACCTCTCTCAATGAATTTCTCGATATACATCTCAACAGACCAA 845
Db      2039 -----ThrProSerGluSerSerSerGluAla-----ProLeuThrSerSerPr 2054
QY      846 GATGCTACACAGAAAGAGAAACAGAAATTCGCTGTG-----TTGAGAGA 890
Db      2054 oAlaThrThrThrGluValileThrGluSerValLySerThrThrProlyGluGl 2074
QY      891 GCAGAGGGTGGAGCTCAGCGTCTCTCTGTTAAACCAGAAATTCAGGAGAGAGCTCGCTGA 950
Db      2074 uSerSerSerGluileThrVallyLeu-----Se 2084
QY      951 CTCACGATCCCATATTTACCAGAGCTAGCAGGAAGTCC-----CAACT 995
Db      2084 rSerlySerProGluValThrGluSerSerVallySerSerProSerThrProSerTh 2104
QY      996 TCAGATGCAAGAAATATTTAAGAAACTTCCAGGATTCAAAAAATTCATCTGTAGGATT 1055

```

```

Db      2104 rThrSerGlnSerValThrSerThrValProGluThrSerLySerThrValLeuSerSe 2124
QY      1056 TAGA-----CCAAAGAAAAGAAAGATGGCTCAAGCTCCACAGAGAT 1097
Db      2124 rGluAlaProValThrSerThrSerProThrGluValHisThrSerSerGluThrlySer 2144
QY      1098 GCAACTTTACGGCATCTTTAAGAGACACAGTGCAGAGCAAAAGCCCTCAAGTCACT 1157
Db      2144 oSerLeuSerAlaSerSerThrThrGlyAspThrAsnSerThrThrProSerSerSe 2164
QY      1158 CCTGTCTTTTGAATCCCAACAAATTCAGAAAGTGAAGTCTATCATGACCACTGAGGA 1217
Db      2164 rLeuAla-----SerVallySerThrSerAlaProGluGlyThrSerAlaSe 2180
QY      1218 GGCAAGCACACGAAATCTATCTCACAGCTCACAGCTCAAAAGCGCTGATCAGCAAGC 1277
Db      2180 rValAlaProVallyLeuSerSerLeuSerProAspValSerGlnProSerThrlysth 2200
QY      1278 ACTAGAGGAGAAACAATCTTTGGATGGGACAAATTCAGTTCATCATGAATTCCTCG 1337
Db      2200 rPheAlaThrGluSer-----SerThrValGlnAlaSerGluThrSerSerGl 2217
QY      1338 A----- 1338
Db      2217 yThrSerVallySerThrSerGluProGluSerHisValThrlySerLeuSerlleThrSe 2237
QY      1339 -----TCCTCGCAGCCTTTGGTCTGCACACCCCAATCAGAGCTGCCAC 1382
Db      2237 rSerAsnProSerSerSerValProValThrSerProlySerThrProThrValProGl 2257
QY      1383 ATCTTTTCTGCTTATTAACAGAGATGCTACTTTGAGTCCAGAACTTCTCTGTTGAACC 1442
Db      2257 uSerThrGluGlnProThrSerThrThrProSerGlyGlnSerLeuThrProMetAsnSe 2277
QY      1443 CCAGCTTGACAGATGGAGGAGAGAG-----CATGTCTA-----CTGA 1484
Db      2277 rAsnSerGluValLeuThrThrSerGluProHisVallySerSerSerLeuSerProAs 2297
QY      1485 CACTTCTTGTCTCT-----CCACCTGCTATGGCCTTCTACCTCCCTGTCAAGAGCTCC 1535
Db      2297 vAlSerGlnSerSerThrThrProAsnAsnLeuSerGluSerSerThrValGluThrPr 2317
QY      1536 ACCTTCTTATGGCATCAAGCATCTTCTCTCTG-----ACTGA 1574
Db      2317 o-----LySerSerGluValSerLeuAsnSerGluGluProSerThrThrGl 2334
QY      1575 TCAGAGCACACACA-----GATACAATGGCCATCACCAGACAAATCTAGTACCAGG 1625
Db      2334 uAlaProThrThrLeuSerProAspIleLeuSerThrThrAsnAsnLeuSerGlnSe 2354
QY      1626 GCTCACCATCCCCACAGTGAATTTCTGCAATCAGCCAA----- 1665
Db      2354 rSerThrValSerThrGluAspArgSerGluileSerSerGluAsnSerGluLyProTh 2374
QY      1666 -----CTGGCTCTGGGAATTTTCATCTCCACCTGCATCTTCA----- 1701
Db      2374 rSerAlaProGluLeuValThrSerSerValThrHisAlaSerSerSerProAspVa 2394
QY      1702 -----GATCAGACCGCATCAAGTGCAGGTGGCGAAGATATGGTCAG 1742
Db      2394 lProThrGluSerSerGluProAspLeuThrGlySerSerThrGluAsnIleProGl 2414
QY      1743 ACACCTAGATGAATGATCTCTCTGACACCTCCGCCCA----- 1782
Db      2414 uAlaSerSerlyGlnThrIleSerSerThrProThrProAspThrThrAlaSerGl 2434
QY      1782 ----- 1782
Db      2434 uGluProThrlySerThrSerMetSerProAspLeuSerThrThrSerAsnValLeuSe 2454
QY      1782 ----- 1782
Db      2454 rGluSerSerThrThrProGluSerSerSerLySerSerProValSerSerSerThrGluGl 2474

```

QY 1783 -----TCTGAGGTACCAGAGCTCAGCGAATATGTTTC 1814  
 Db 2474 yIleSerValValThrSerThrGluPheSerLysValProGlu---SerThrIleSerSe 2493  
 QY 1815 TGTCCCGAGATCTTCTGAGGATACCACTCTCTCTCAGCTTTTACAGATATACACAC 1874  
 Db 2493 rValLeuGluGluAspLeuThrLysThrProSerProIleLeuGluGluThrThr 2513  
 QY 1875 TAGTTCTATGACCATTTGCCCAAGGCCGAGAGCTGTAGTGTCTTCTCAGTCTGCGTGT 1934  
 Db 2513 rAlaSerGluThrSerGluProLeuThrGluAspSerLeuThr---ValSerValArgII 2532  
 QY 1935 TGCTAACATGGCTTCTCAACGACCTGTTCACAGAGCTCT 1977  
 Db 2532 eHieGluLeuThrThrSerSerGluAenValProLysGluSer 2546  
 RESULT 3  
 A82255  
 hypothetical protein VC0998 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
 C:Species: Vibrio cholerae  
 C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
 C/Accession: A82255  
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
 chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, R.  
 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 406, 477-483, 2000  
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
 A:Reference number: A82035, MUID:20406833; PMID:10952301  
 A:Accession: A82255  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1621 <HEI>  
 A:Cross-references: GB:AE004181; GB:AE003852; NID:G9655454; PIDN:AAF94159.1; GSPDB:GN001  
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C:Genetics:  
 A:Gene: VC0998  
 A:Map position: 1  
 Alignment Scores:  
 Pred. No.: 0.0232 Length: 1621  
 Score: 157.00 Matches: 126  
 Percent Similarity: 35.98% Conservative: 109  
 Best Local Similarity: 19.24% Mismatches: 236  
 Query Match: 2.67% Indels: 184  
 DB: 2 Gaps: 30  
 US-10-007-270-1 (1-3330) x A82255 (1-1621)  
 QY 736 TCAGAGACAAATTCAGATGTTGCCAACGTCTCAGTGGCGCTTTCCCTCTCACTCT 795  
 Db 827 AlaAspGluLeuLeuAlaAspLeuAlaGlnProGlnSerAsnThrValAspThrSer 846  
 QY 796 GATGACACCTCTCAATGAATTCCTCGATAAT-----ACACTCAAC 837  
 Db 847 AspAlaLeuAlaProAspGlyLeuSerGlnSerValGluGluProLeuThrLeuAsn 866  
 QY 838 GACACCAAGTCTCTACACAGAAAGAACAGATTCGCT----- 879  
 Db 867 AspLeuGluLeuProGluGluAenAspGluProGlnLeuAlaGluValThrProSerSer 886  
 QY 880 GTGTTGAGGACAGAGGGAGCTCAGCTCTCTCTGTTAAACACAGAGTTCAAGGCA 939  
 Db 887 AlaPheAspGluGlnValGluThrGluLeuLeuProGluSerGluProLeuAlaAla 906  
 QY 940 GAGTCGCTGACTCCCACTCC-----CCATATTAC 969  
 Db 907 GluAlaSerAenAspGluSerAspLeuThrAlaLeuAsnGluLeuAspLeuProGluThr 926  
 QY 970 CAG-----GAGTACGAGAAAGTCCCACTCAG 999  
 Db 927 ThrGluGluAspValLeuAlaAspValGlnLeuGluProAlaAlaGluSerGluValGlu 946

QY 1000 ATGCAAAAGATATTAAAGAACTTCCA-----GGATTCAAAAATAATCCATGTG 1047  
 Db 947 ProAspLeuGluLeuValAsnGluProValThrGluGluAlaPheThrGluLeuAspGlu 966  
 QY 1048 TTAGGATTTAGACAAAGAAAGATGGCTCAAGCTCCACAGAGATGCAACTTACG 1107  
 Db 967 LeuAspLeuProGluThrGluGluAspAlaLeuAlaAspAlaGlnLeuGluProVal 986  
 QY 1108 GCATCTTTAAGACACACAGTGCAGAA-----GCAAAAGCCCTCAAGT 1152  
 Db 987 Ala-----GluSerGluValGluProGluLeuAspLeuAlaSerGluProAlaGlu 1003  
 QY 1153 GACTCTCTCTTTGATTCCACAAATTTGAAAGTGAGAGAGTCTATCATCGAACCATG 1212  
 Db 1004 GluGluAlaPheThrGluLeuAsnLysLeuAspLeuProGluThr-----Thr 1019  
 QY 1213 GAGGAGGACCAAGCAACCAATCTATCTC---ACAGTACAGACCTCAAA----- 1260  
 Db 1020 GluGluAspAlaLeuAlaAspAlaGlnLeuGluSerAlaThrGluSerGluValGluSer 1039  
 QY 1261 -----AGGCTGATCAGCAAGACACTAGAGGAGCAACAATCTTTGATGTGGGCAATT 1314  
 Db 1040 GluLeuGluLeuValSerGluProAlaAlaGluAlaPheThrGluLeuAspGluLeu 1059  
 QY 1315 -----CAGTTCACTGATGAATTTGCTGATCACTGCCAGCCTTTGCTCTGACACC 1365  
 Db 1060 AspLeuProGluThrGluGluAspAlaLeuAlaAspSerGlnLeuGluProAlaAla 1079  
 QY 1366 CAACTCAGAGCTGCCCACTCTTTTGTGTTATAACA----- 1401  
 Db 1080 GluSerGluValGluProGluLeuLeuValSerGluProValThrGluGluAlaPhe 1099  
 QY 1402 -----GAGGATGCTACTTTGAGTCCA 1422  
 Db 1100 ThrGluLeuAspGluLeuAspLeuProGluThrGluGluAspAlaLeuAlaAspAla 1119  
 QY 1423 GAATCTCTCTCT-----GTTGAACCCAGCTTGAGACAGTGGACGAGCA 1467  
 Db 1120 GlnLeuGluProAlaValGluSerGluValGluProGluLeuGluLeuAla----- 1136  
 QY 1468 GAGCATGCTCTACCTGACACTTTGTTGTTCCACTCTCT-----ATGCGCTCTACTCTCC 1521  
 Db 1137 -----SerGluProAlaGluGluAlaSerThrGlu 1147  
 QY 1522 CTGTCAAGAGCTCCACTTTCTTTATGGCATCAGCATCTCTCTCTGACTGATCAAGGC 1581  
 Db 1148 LeuAsnGlu-----LeuAspLeuProGluThr 1157  
 QY 1582 ACCACAGATACAATGGCCACTGACACCAATGTAGTACCAGGCTCACCATCCCAACC 1641  
 Db 1158 GluGluAspAlaLeuAlaAspAlaGlnLeuGlu-----ProAla 1170  
 QY 1642 AGTGATTTATTCGAATCAGCCCACTGGCTCTGGGA----- 1677  
 Db 1171 AlaGluSerGluValGluSerGluLeuGluLeuAlaSerAspLeuGluGluGluPro 1190  
 QY 1678 -----ATTTCACATCCACTCTTCAGATGATGACAGCGCATCAAGT 1719  
 Db 1191 PheThrGluLeuAsnGluLeuAspLeuProGluThrGluGluAspAlaLeuAlaAsp 1210  
 QY 1720 GCAGGTGGCGGAGATATGGTCAGACAC-----CTAGATGAATGGATCTCTGTGACACT 1773  
 Db 1211 AlaGlnLeuGluProAlaValGluSerGluValGluProGluLeuGluLeuAlaThrGlu 1230  
 QY 1774 CTTGCCCCCATCTGAGTA---CCAGAGCTCAGCAATATGTTCTGTCGCCAGATCAATTTC 1830  
 Db 1231 ProAlaGluGluGluAlaPheProGluLeuAspGlu---LeuAspLeuProGluThr 1249  
 QY 1831 TTGGAGATACCACTCTCTCTCAGCTTTACAGTATATCACCCTAGTCTCTATGACCAATT 1890  
 Db 1250 GluGluAspAlaLeuAlaAspAlaGlnLeuGlu-----ProAlaAlaGluAlaGluVal 1267  
 QY 1891 GCCCCCAAGGCCGAGAGCTGGTGTCTTCAGTCTGCTGTGCTTAACATGGCCTTC 1950

```

Db      1268 GluProGlu-----LeuGluLeu 1273
          |||:::
QY      1951 TCAAGACGCTGTTCAACAGAGCTCTCTGAGTACGAGCTCTGGAGCAACAATTCACA 2010
          :::::|||||
Db      1274 AlaseAspLeuGluGluys-----GluThrPheThr 1284
          :::::
QY      2011 CAGCTGCTGGTTCATATCTACGATCCAAATCTTACAGGATTAAAGCAACTTGAATACTT 2070
          :::::|||||
Db      1285 GluLeu-----AspGluLeuAspLeuPro 1292
          :::::
QY      2071 AACTTCAGAACGGAGTGTGATGTGAATGACAAATGAAGTTTGTCTAGCTGTGCGG 2130
          :::::
Db      1293 GluTyThrGluGluAspAlaLeuAlaGlnLeuGluProAlaAlaGluSerGlu 1312
          :::::
QY      2131 TATACCTCACAAGCTGTGCACGGGCTCTTGAGGATTTTCTGCTGCGAGCCCAA 2190
          :::::
Db      1313 ValGluProGluLeuGluLeuAlaSerValLeuGluGlu-----GluAspSerPheThrGlu 1331
          :::::
QY      2191 CAATCCATCTCGAATACAGACGCTACTCT-----CTCAACATT 2229
          |||:::
Db      1332 LeuAspGluLeuAspLeuProGluTyThrGluLysAspAlaLeuAlaAspAlaGlnLeu 1351
          :::::
QY      2230 GAACAGCT-----GATCAAGCAGATCCCTGCAAGTTCTCTGGCTGCGCGCAATTGCC 2283
          |||:::
Db      1352 GluProAlaAlaGluSerGluValGluPro-----GluLeuGluAspGluThr 1367
          :::::
QY      2284 CAATGTGTAAGAACGACGAGCTGAGGAGCGAGTGTGCTGCAACACGAGATATGAC 2343
          :::::
Db      1368 GluThrLeuAlaGlnGluThrGluSerAspAlaLeuValAlaAspGluAspLeuLeuAla 1387
          :::::
QY      2344 AGC---CAGGGAGCTGCGAGCTGTGGAACGAGCTCTGTGGC 2385
          :::::
Db      1388 SerValGluSerAlaValAspGluValGlnProGluLeuGlu 1402
          :::::

```

## RESULT 4

```

H81236
DNA topoisomerase I NMB0118 [imported] - Neisseria meningitidis (strain MC58 serogroup B
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: H81236
R;Rettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haf, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Mabinani, V.; Pizzi, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307
A;Accession: H81236
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-768 <TET>
A;Cross-references: GB:AE002370; GB:AE002098; NID:g7225327; PIDN:AAF40577.1; PID:g722533
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB0118
C;Superfamily: DNA topoisomerase I

```

## Alignment Scores:

```

Pred. No.:      0.0242      Length:      768
Score:          155.50      Matches:     140
Percent Similarity: 33.54%      Conservative: 75
Best Local Similarity: 21.84%      Mismatches: 243
Query Match:      2.65%      Indels:    183
DB:              2          Gaps:      30

```

US-10-007-270-1 (1-3330) x H81236 (1-768)

```

QY      859 GAAGAGAAACAGAAATTCGCTGTGTGAGAGAGAGAGG-----GTGAGCTCAGC 909
          |||:::
Db      181 GluArgGluAenGluLeuArgAlaPheGluAlaGlnGluTyTrpThrValHisLeuAsp 200
          :::::
QY      910 GTCTCTCTGGTAAACCAAGAGTTCAAGCGAGAGCTCGCTGACTCCAGTCCCATATTAC 969
          :::::

```

```

Db      201 SerHisGlyArgSerLysPheThrAlaLysLeuAlaGlnTyAsnGlyAlaLysLeu 220
          |||:::
QY      970 CAGAGCTAGCAGGAAGTCCCACTTCAGATGCAAAAGATATTTAAGAAATTTCCAGGA 1029
          :::::
Db      221 GluGlnPheAspLeuProAenGluAlaAlaGlnAlaAspValLeuLysGluLeuGluGly 240
          :::::
QY      1030 TTCAAAAAATCCATGTGTAGGATTAGACCAAGAAGAAAAAGATGGCTCAAGCTCC 1089
          :::::
Db      240 ----- 240
QY      1090 ACAGAGATCAACTTACGGCATCTTTAAGAGACAGATGCAGAGCAAGCAAAAGCCTGCA 1149
          |||:::
Db      241 LysGluAlaValValThrAlaLalGluLysLys-----LysArgSerArgAsnProAla 258
          :::::
QY      1150 AGTGACCTCTGTCTTTTGATTCCAACAAAATTGAAGTGAGGAAGTCTATCATGAACC 1209
          :::::
Db      259 AlaProPheThrThr-----SerThr 265
          :::::
QY      1210 ATGGAGGAGGAGCAAGCAACCAAGATCTATCTCACAGTACAGACCTCAAAAGCTGATC 1269
          |||:::
Db      266 MetGlnAlaValValThrAlaLalGluLysLeuGlyPheThr-----ArgThrMet 282
          :::::
QY      1270 AGCAAGCAGCTAGAGGAGCAACAACTCTTTGGATGTGGG----- 1308
          |||:::
Db      283 ArgThrAlaGlnGlnLeuTyGluGlyLeaSpValGlyGlnGlyAlaLalGlyLeuLe 302
          :::::
QY      1309 ACATTCAGTTCATCATGATCAAAATTCGTGGATCTACCTGCCAGCCTTTGGTCTCGACACCCA 1368
          |||:::
Db      303 ThrTyMetArgThrAspSerVal-----AsnLeuAla 313
          :::::
QY      1369 TCAGAGCTGCCACATCTTTGCTGTGTATTAACAGAGATGTCTTTGAGTCCAGAACTT 1428
          |||:::
Db      314 AspGluAlaLeuThrGluLeuArgHisTyrlleGluAsnLysLllegLysGluTyLeu 333
          :::::
QY      1429 CCTCTGTGTGAACCCAGCTTGACAGATGGACGGAGCAGAGCATGTCTACCTACCTACACT 1488
          |||:::
Db      334 ProSerAlaAlaLysGlnTyLysThrLysSer-----LysAsnAla 347
          :::::
QY      1489 TCTTGCTCTCCACTGCTATGGCTCTACCTCCCTGTGTCAGAAGCTCCA----- 1536
          :::::
Db      348 GlnGluAlaHisGluAlaIleArgProThrSerValTyArgThrProGluSerValLys 367
          :::::
QY      1537 CTTTCTTTTATGGCATCA-----AGCATCTCTCTGACTGATCAAGGCAACACAGAT 1590
          |||:::
Db      368 ProPheLeuSerAlaAspGlnPheLysLeuTyGlnMetIleTrpGlnArgThrValAla 387
          :::::
QY      1591 ACAAATGGCCACT-----GACCAGACAATGCTAGTACCGGGCTACCATCCCC 1638
          |||:::
Db      388 CysGlnMetThrProAlaLysPheAspGlnThrVal-----AspIleThrValGly 405
          :::::
QY      1639 ACCAGTCAATATTCGCAATCAGCCAACTG-----GCTCTGGAAATTTTCATCCA 1689
          :::::
Db      406 LysGlyValPheArgValThrGlyGlnValGlnThrPheAlaGlyPheLeuSerValTy 425
          :::::
QY      1690 CTTGCATCTTCAGATGACAGCCGATCAAGTGCAGTGGCGGAGAGATATGTCAGACACCTA 1749
          |||:::
Db      426 GluGluSerSerAspAspGluGlu-----GlyGluAsp----- 436
          :::::
QY      1750 GATGAAATGATCTGTCTGACACTCTCTGCCCATCTCAGGTACCAAGCTCAGCGAATAT 1809
          :::::
Db      437 -----SerLysLysLeuProGluMetSerGlu----- 445
          :::::
QY      1810 GTTCTCTGCCAGATCATTTCTTGAGGATACCATCTCTGTCTCAGCTTTA----- 1860
          |||:::
Db      446 -----GlyAspLysLeuProValAspLysLeuTyGlyGlu 457
          :::::
QY      1861 CAGTATATCACCATCTAGTCTCTATGACCATTCGCCCAAGGGCCGAGAGCTGGTAGTGTTC 1920
          |||:::
Db      458 GlnHisPheThrThrPro-----ProProArgTyAsnGluAlaThrLeuVal 473
          :::::
QY      1921 TTCAGTCTGCGTGTGTGCTAACATG-----GCCTTCTCCACAGCCTGTTTCAAC 1968
          :::::

```





type III secretion protein SctC CP0044 [imported] - Chlamydophila pneumoniae (strains CW  
C/Species: Chlamydophila pneumoniae, Chlamydia pneumoniae

C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000

C;Accession: F72045; G81618  
R;Kalkan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21, 385-389, 1999  
A>Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A;Reference number: A72000; MUID:99206606; PMID:10192388  
A;Accession: F72045  
A:Molecule type: DNA  
A:Residues: 1-919 <ARN>  
A:CROSS-references: GB:AE001652; GB:AE001363; NID:g4376997; PIDN:AAD18841.1; PID:g4377000  
A:Experimental source: strain CWL029  
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A>Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A;Reference number: AB1500; MUID:20150255; PMID:10684935  
A;Accession: G81618  
A:Molecule type: DNA  
A:Residues: 1-919 <REM>  
A:CROSS-references: GB:AE002168; GB:AB002161; NID:g7188982; PIDN:AAF37937.1; PID:g7188989  
A:Experimental source: strain AR39, HL cells  
C;Genetics:  
A;Gene: yscC; CP0044

Alignment Scores:

| Pred. No.:             | Score: | Length:          |
|------------------------|--------|------------------|
| 307                    | 153.00 | Matched: 919     |
| Percent Similarity:    | 34.79% | Mismatches: 133  |
| Best Local Similarity: | 20.03% | Conservative: 98 |
| Query Match:           | 2.60%  | Indels: 246      |
| DB:                    | 2      | Gaps: 187        |

US-10-007-270-1 (1-3330) x F72045 (1-919)

| QY  | DQ   | Score                  | Length           |
|---|------|------------------------|------------------|
| 438 AGGAAGACGATGGGAAGCATATCGATCTTTTCGTGATCGATCCCTGCACAGGGGAAT       | 497  | 153.00                 | Matched: 919     |
| 20 lysylslyllecyllyleLeuSerGlyLeuPhePheLeuAspLeuValLeuGluVal        | 39   | Percent Similarity:    | Mismatches: 133  |
| 498 ATCAGGACTGGTGCAGCATCTGCCAGCAGGAGACCTTCCTGCCTCTTGACATTGGCAAAA    | 557  | Best Local Similarity: | Conservative: 98 |
| 40 SerSerGlnArgProThrGluThrSerAlaAsnVallyshisasnLeuArgaspGluLys     | 59   | Query Match:           | 2.60%            |
| 558 ACTTCAGCAATTCACAGGACGACCTGATCTTCCTCCAGCAGAGATAAAACAGAGAAGTT     | 617  | DB:                    | 2                |
| 60 LeuAlaAlaCysProLys-----AsnSerAlaAla                              | 69   |                        |                  |
| 618 TCCTTCACAGAAAGATCAATATCTGCAGAGAGACATTTGGGAGGCGCTGGTAACCA        | 677  |                        |                  |
| 70 SerLeuSerAlaLysLysSerHisThrLysLysThrProGlySerIle-----            | 86   |                        |                  |
| 678 TTGTCAATTTCAACAGCAATCTACAATTTCAAAGACTTCGGCAGATATCTAACAGAAACCCTC | 737  |                        |                  |
| 87 -----ProSerLysValPheSerLysPheAspAlaThrGlnAspLysThrPhe            | 102  |                        |                  |
| 738 AGAAGACGAATTCAGATTTGCCACGCTCTCACTTGGGCGCTTTCCCTCACTCCCTCGA      | 797  |                        |                  |
| 103 GlnLysThrSerglyser-----AlaPheProAla-LysPro--                    | 114  |                        |                  |
| 798 TGACACCTCTCTCAATCAATTTTCGATAATACACTCAACGACACCAAGATGCCTACAAC     | 857  |                        |                  |
| 115 ----ThrThrLeuLysGlu-----LeuGluGluArgLysLysProArgPr              | 128  |                        |                  |
| 858 AGAAGAGAAACAGAAATTCGCTGTG-----TTGAGGAGCAGAGGCT                  | 899  |                        |                  |
| 128 oGluArgargThrThraAspValLysArgSerProArgPheLeuProThrGlnGluVa      | 148  |                        |                  |
| 900 GGAGCTCAGCTCTCTCTGTTAAACCAAGATTTCAAGGACGAGCTCGCTCACTCCCAGTC     | 959  |                        |                  |
| 148 lGluGluProVal-----  | 152  |                        |                  |
| 960 CCATATTACAGGAGCTAGCAGGAAAGTCCCACTTCAGATGCAAAAGATATTAAAG         | 1019 |                        |                  |
| 153 -----ProAlaAlaSerLysGluGlnLeuAspSerIleGlnValTrpGluGl            | 168  |                        |                  |

1968 CAAAGAGCTCTCTGGAGTACCGAGCTCTGGAGCAACAAATTCACACAGCTCTGGTTCATA 2027  
Db 481 |||||  
-----GluGlnLeuLeuLysSerLeuAspValProG1 491  
QY 2028 TCTGATGACCAATCTT----- 2043  
Db 491 uMetAlaHisThrIleuAspPProAlaSerThrAlaLeuAlaLeuGlyGlyThrGlyTh 511  
QY 2044 -ACAGATTAAACAACCTTGAATA-----CTTAACTTCAGAAACGGGAGTGT 2090  
Db 511 rThrSerProLysSerLeuAluPhePheMetTyrLysLeuLysTyrGlnAsnGlyGluVa 531  
QY 2091 GATTCTGAATGACAAAAGAAATTTGCTAGTCTCTGCGGTATACCTC-----ACCAA 2144  
Db 531 lIleAlaAsnAla-----LeuGlnAspIleGlyTyrAsnLeuTyrValThrTh 547  
QY 2145 GCGTGTGACGGGGTCTTGAGGATTTCGTTGCTGCGAGCCCAACACACTCCACTCTGGA 2204  
Db 547 xAlaMetAsp-----GluAspPheIleAsnThrIleuAsnSerIleGlnTrpLeuG1 564  
QY 2205 AATAGACAGC 2214  
Db 564 uValAsnAsn 567  
RESULT 7  
S28916  
N:Alternate names: duchenne muscular dystrophy protein  
C:Species: Mus musculus (house mouse)  
C:Date: 22-Nov-1993 #sequence revision 03-Nov-1995 #text\_change 16-Jul-1999  
C:Accession: S28916; B27162; S10922; C43837; B40134  
R:R,Bies, R.D.; Phelps, S.F.; Cortez, M.D.; Roberts, R.; Caskey, C.T.; Chamberlain, J.S.  
Nucleic Acids Res. 20, 1725-1731, 1992  
A:Title: Human and murine dystrophin mRNA transcripts are differentially expressed during  
A:Reference number: S28916; MUID:92253376; PMID:1579466  
A:Accession: S28916  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-3678 <BIE>  
A:Cross-references: EMBL:M68859  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991  
M. Hoffman, E.P.; Bertelson, C.J.; Monaco, A.P.; Feener, C.; Kunkel, L.M.  
Cell 50, 509-517, 1987  
A:Title: Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary  
A:Reference number: A90897; MUID:87273512; PMID:3607877  
A:Accession: B27162  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-201 <KOE>  
R:Nudel, U.; Zuk, D.; Einat, P.; Zeelon, E.; Levy, Z.; Neuman, S.; Yaffe, D.  
Nature 337, 76-78, 1989  
A:Title: Duchenne muscular dystrophy gene product is not identical in muscle and brain.  
A:Reference number: S06461; MUID:89082658; PMID:2909892  
A:Accession: S10922  
A:Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-106 <NUD>  
A:Cross-references: EMBL:X14183  
R:Rapaport, D.; Lederfein, D.; den Dunnen, J.T.; Grootsholten, P.M.; Van Ommen, G.J.; F.  
differentiation 49, 187-193, 1992  
A:Title: Characterization and cell type distribution of a novel, major transcript of the  
A:Reference number: A43837; MUID:92316332; PMID:1377655  
A:Accession: C43837  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 'MREHLXG', 3069-3181 <RAP>  
A:Note: sequence extracted from NCBI backbone  
Hoffman, E.P.; Monaco, A.P.; Feener, C.C.; Kunkel, L.M.  
Science 238, 347-350, 1987  
A:Title: Conservation of the Duchenne muscular dystrophy gene in mice and humans.  
A:Reference number: A40134; MUID:89018015; PMID:3659917  
A:Accession: B40134  
A:Status: preliminary

Db 2176 SerLysThrAspValAsnIleLeuGlnGluLysLeuSerLeuArgTrpHis 2195  
 QY 724 ----CTAAGAAACCCCTCAGAAGACAA-----747  
 Db 2196 AspIleCysLysGluLeuAlaGluArgLysArgIleGluGlnLysAsnValLeu 2215  
 QY 748 -----ATTCAAGATGTTGCCACGTC 768  
 Db 2216 SerGlnPheGlnArgAspLeuAsnGluPheValLeuTrpLeuGluIleAlaAspAenIle 2235  
 QY 769 TCACCTGGGCGTTCCCTCTCACTCTCGATGATCACCTCTCAATCGAATTCTCGATAT 828  
 Db 2236 AlaIleThrProLeu-----GlyAspGluGlnLeuLysGlu-----2248  
 QY 829 ACATCTCAACGACACCAAGATGCCCTACACAGAGAGAAACAAGATTGCTGTGTGGAG 888  
 Db 2249 GlnLeuGluGlnValLysLeuLeuAlaGluGluLeuProLeuArgGlnGlyIleLeuLys 2268  
 QY 889 GAGCAGAGGGTGGAGCTCAGCGTCTCTGTGTGAACACAGAGTTCAAGCGAGAGTCGCT 948  
 Db 2269 GlnLeuAsnGluThrGlyGlyAlaValLeuValSerAlaProIleArgProGluGln 2288  
 QY 949 GACTCCCGCTCCCATATTACGAGAGCTAGCAGGAAGTCCCACTTCAGATGCAAG 1008  
 Db 2289 AspLys-----LeuGluLysLysLeuLysGlnThrAsnLeuGlnTrpIleLys 2304  
 QY 1009 ATATTAAAGAACTCCA-----GGATTCAAAAAATCCATGTGTTAGGATTAGA 1059  
 Db 2305 ValSerArgAlaLeuProGluLysGlnGlyLeuGluValHisLeuLysAspPheArg 2324  
 QY 1060 CCAAGAGAGAAAAGATGGCTCAAGCTCCACAGAGATCAACTCAGGCATCTTTAAG 1119  
 Db 2325 GlnLeuGluGln-----LeuAspHisLeuLeuLeuTrpValSerProIle 2340  
 QY 1120 AGCACAGTGCAGAGCAAAAGCCCTGCAAGTGACTCTGTTGATTGATCCAAACAA 1179  
 Db 2341 ArgAsnGlnLeuGluIleTyrAsnGlnProSerGlnAlaGlyProPheAspIleLysGlu 2360  
 QY 1180 ATTGAAGTGAAGAGTCTATCATGGAACCATGGAGAGGACAAGCAACACAGAAATCTAT 1239  
 Db 2361 Ile-----GluValThrValHisGly-----LysGln-----2369  
 QY 1240 CTCACAGCTACAGACCTCAAAAGGCTGATCAGCAAGCA-----CTAGAGAGAGACAA 1293  
 Db 2370 -----AlaAspValGluArgLeuSerLysGlyGlnHisLeuTyrLysGluLys 2386  
 QY 1294 TCTTTGGATGTGGGCAATTCACTTCATGATGAATTCGTGGATCACTGCCAGCCTTT 1353  
 Db 2386 -----2386  
 QY 1354 GGTCCTGACACCAATCAGAGTCCCAATCTTTGCTGTTATACAGAGGATGCTACT 1413  
 Db 2387 ---ProSerThrGln-----2390  
 QY 1414 TTGAGTCCAGACTTCCTCTGTGTGAACCCACGCTTCGAGACAGTG-----GAC 1461  
 Db 2391 -----ProValLysArgLysLeuGluAspLeuArgSerGluTrpGlu 2404  
 QY 1462 GGAGCAGACATGGTCTA-----CCTGACACTCTTGTGCTCCCA 1500  
 Db 2405 AlaValAsnHisLeuLeuArgGluLeuArgThrLysGlnProAspArgAla-----2421  
 QY 1501 CCTGCTATGGCCTCTACCTCCCTGTCAGAGGCTCCACTTCTTTATGGCATCAAGATC 1560  
 Db 2422 ProGlyLeuSerThrThrGlyAlaSer-----AlaSerGlnThr 2434  
 QY 1561 TTCTCTCTGACTGATCAAGGCACCAAGATACATGGCCACTGACCAAGCAATGTAGTA 1620  
 Db 2435 ValThrLeu-----ValThrGlnSerValValThrLysGluThr-----ValIle 2449  
 QY 1621 CCAGGGCTCACCTCCGACAGTGATTTCTGCAATTCAGCCAACTGGGCTCTGGGAATT 1680

Db 2450 SerLysLeuGluMetProSerSer-----LeuLeuLeuGluVal 2462  
QY 1681 TCACATCCACTGTCATCTTCAGATGACAGCCGATCAAGTCAGGTGGCGAAGATATGTC 1740  
Db 2463 -----ProAlaLeuAlaAspPheAsnArgAlaTrpThrGluLeuThrAspTrpLeu 2479  
QY 1741 AGACACTAGATGAATGGATCTGCTGACACTCTCCGCCATCTGAGGTACACAGATC 1800  
Db 2480 SerLeuLeuAspArgValIleLysSerGlnArgValMetValGlyAspLeuGluAspIle 2499  
QY 1801 AGCGAATATGTT-----TCTGTCCCATCATCTTCTTCGGAGTACCACCT 1845  
Db 2500 AsnGluMetIleIleLysGlnLysAlaThrLeuGlnAsp-----LeuGluGlnArg 2517  
QY 1846 CCTGTCTCAGCTTACAGTATATCACCACTAGTTCTATGACCAATGCCCCCAAG----- 1899  
Db 2518 Pro-----GlnLeuGluGluLeuIleThrAlaAlaAsnLeuLysAsnLysThrSer 2535  
QY 1900 GCGCCGAGAGCTGTAGTGTCTTCAGCTCGGTGTCTTAACATGCGCTTCTCCAAAGAC 1959  
Db 2536 AsnGlnGluAlaArgThrIleIleThrAspArgIleGluArgIleGlnIleGlnTrpAsp 2555  
QY 1960 -----CTGTTCAAC-----AAGAGCTCT 1977  
Db 2556 GluValGlnGluGlnLeuGlnAsnArgArgGlnGlnLeuAsnGluMetLeuLysAspSer 2575  
QY 1978 CTGGAGTACCAGCTCTGGAGCAACAATTACACAGCTGCTG----- 2019  
Db 2576 ThrGlnTrpLeuGluAlaLysGluGluAlaGlnValIleGlnValArgGlyLys 2595  
QY 2020 -----GTTCCATATCTACGATCCATCTTACAGA 2049  
Db 2596 LeuAspSerTrpLysGluGlyProHisThrValAspAlaIleGlnLysLysIleThrGlu 2615  
QY 2050 TTTAGCACTT---GAAATACTTAACCTCAGAAACGGAGTGTGATTGTG-----AAT 2100  
Db 2616 ThrLysGlnLeuAlaLysAspLeuArgGlnArgGlnIleSerValAspValAlaAsnAsp 2635  
QY 2101 AGCAAAATGAAGTTCTGAAGTCTGTGCGCTATACCTCACCAAGGCTGTGACGGGTC 2160  
Db 2636 LeuAlaLeuLysLeuLeuAspTyrSerAlaAspThrArgLysValHisMetIle 2655  
QY 2161 TTGGAGGATTT-----CGTCTGCT 2181  
Db 2656 ThrGluAsnIleAsnThrSerTrpGlyAsnIleHisLysArgValSerGluGlnGluAla 2675  
QY 2182 GCAGCCCAACATCCATCTCGAATAAGACAGCTACTCTCTCAACATTTGAACAGCTGAT 2241  
Db 2676 AlaLeuGluThrHisArgLeuLeuGlnGlnPheProLeuAspLeuGlu----- 2692  
QY 2242 CAAGCAGATCCCTGCAAGTCTCGCTCGCGGAGATTGGCCCAATGTGTAAAGAAC--- 2298  
Db 2693 -----LysPheLeuSerTrpIleThrGluAlaGluThrAlaAsnVal 2707  
QY 2299 -----GAACGAGCTGAGGAAGCGAGTGTGCG 2325  
Db 2708 LeuGlnAspAlaSerArgLysGluLysLeuLeuGluAspSerArgGlyValArgGluLeu 2727  
QY 2326 TGCACACAGGATATGACAGCCAGGGAGCTTGGAC-----GGTCTGGAACACGAGC 2376  
Db 2728 MetLysProTrpGlnAspLeuGlnGlyGluIleGluThrHisThrAspIleTyrHisAsn 2747  
QY 2377 CTCTGTGGCCCTGCGACAAAGGAATCGAGGTCTCTCCAGGGAAGGAGCTCCATGACG 2436  
Db 2748 LeuAspGlnAsnGlnLysIleLeuArgSerLeuGluGlySerAspGluAlaProLeu 2767  
QY 2437 TTGCCAGATCACTCTGAATAATCAAGCATACAAAATAGT-----GTTAAAGTTCCAA 2490  
Db 2768 LeuGlnArgLeuAspAsnMetAsnPheLysTrpSerGlnLeuGlnLysLysSerLeu 2787  
QY 2491 AATCAACAAAATAACAGGTAATCAGTAAAGAAATTTCTGAATTAATCACTGCGGAGGAATAT 2550  
Db 2788 AsnIleArgSerHisLeuGluAlaSerSerAspGlnTrpLysArgLeuHisLeuSerLeu 2807

QY 2551 GAACAA 2556  
Db 2808 GlnGlu 2809  
RESULT 8  
A82009  
DNA topoisomerase (EC 5.99.1.2) I NMA0156 [imported] - Neisseria meningitidis (strain Z2-  
C;Species: Neisseria meningitidis  
C;Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C;Accession: A82009  
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
F.; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A;Reference number: A81775; MUID:2022556; PMID:10761919  
A;Accession: A82009  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-768 <PAR>  
A;Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CA883470.1; PID:g737892  
A;Experimental source: serogroup A, strain Z2491  
C;Genetics:  
A;Gene: topA; NMA0156  
C;Superfamily: DNA topoisomerase I  
C;Keywords: isomerase

Alignment Scores:  
Pred. No.: 0.0382 Length: 768  
Score: 152.50 Matches: 139  
Percent Similarity: 33.54% Conserved: 76  
Best Local Similarity: 21.68% Mismatches: 243  
Query Match: 2.60% Indels: 163  
DB: 2 Gaps: 30

US-10-007-270-1 (1-3330) x A82009 (1-768)

QY 859 GAAAGAGAAACAGAAATTCGCTGTGTGGAGGACGAGG-----GTGGAGCTCAGC 909  
Db 181 GluArgGluAsnGluIleArgAlaPheGluAlaArgGluTrpThrValHisLeuAsp 200  
QY 910 GTCTCTCTGTPAAACCAAGTTCAGGCGAGAGCTCGCTGACTCCAGTCCCATATTAC 969  
Db 201 SerHisLysGlyArgSerLysPheThrAlaLysLeuAlaGlnTyrAsnGlyAlaLysLeu 220  
QY 970 CAGGAGCTAGCAGGAAAGTCCCACTCAGATGCAAAAGATATTTAGAAACTCCAGGA 1029  
Db 221 GluGlnPheAspLeuProAsnGluAlaAlaGlnAlaAspValLeuLysGluLeuGluGly 240  
QY 1030 TTCAAAAAATCCATCTGTGTAGGATTTAGACCACCAAGAAAGAAAGATGGCTCAAGCTCC 1089  
Db 240 ----- 240  
QY 1090 ACAGAGATGCACTTACGCGCATCTTTAAGACACACAGTCGACAGCAAGCAAAAGCCTGCA 1149  
Db 241 LysGluAlaValThrAlaIleGluLysLys-----LysArgSerArgAsnProAla 258  
QY 1150 AGTGACCTCTCTCTTTTGTATTCACAAATTTGAAAGTGGAGGAGTCTATCATGGAACC 1209  
Db 259 AlaProPheThrThr-----SerThr 265  
QY 1210 ATGGAGGAGCAGCAAGCAACCACTATCTCAGCTACAGACCTCAAAAGGCTGATC 1269  
Db 266 MetGlnGlnAspAlaValArgLysLeuGlyPheThr---ThrAsp-----ArgThrMet 282  
QY 1270 AGCAAGCACTAGAGGAAGCAATCTTTGGATGTGGGG----- 1308  
Db 283 ArgThrAlaGlnLeuTyrGluGlyIleAspValGlnGlyAlaIleGlyLeuIle 302  
QY 1309 ACAATTTCAGTTCACTGATGAATTTGCTGGATCATCTGCCAGCTTTGGTCTGACACCA 1368  
Db 303 ThrTyrMetArgThrAspSerVal-----AsnLeuAla 313

QY 1369 TCAGAGCTGCCACATCTTTGCTGTATATACAGAGGATGCTACTTTGAGTCCAGACTT 1428  
 Db |||||  
 QY 314 AspGluAlaLeuThrGluIleArgHisTyrIleGluAsnLysIleGlyLeuTyrLeu 333  
 |||||  
 QY 1429 CCTCTCTTGAACCCACCTTGAGACAGATGGAAGCGAGCAGAGCATGGTCTACCTGACACT 1488  
 Db |||||  
 QY 334 ProSerAlaAlaLysGlnTyrLysThrLysSer-----LysAsnAla 347  
 |||||  
 QY 1489 TCTTGGTCTCCACCTGCTATGGCTCTACCTCTCTGTGAGAAGCTCCA----- 1536  
 Db |||||  
 QY 348 GlnGluAlaHisGluAlaIleArgProThrSerValTyrArgThrProGluSerValLys 367  
 |||||  
 QY 1537 CCTTCTTTATGGATCA-----AGCATCTCTCTCTGACTGATCAGGACGACACAGAT 1590  
 Db |||||  
 QY 368 ProPheLeuSerAlaAspGlnPheLysLeuTyrGlnMetIleThrGlnArgThrValAla 387  
 |||||  
 QY 1591 ACAATGCCACT-----GACCAGACAATGCTAGTACCAAGGCTCACCATCC 1638  
 Db |||||  
 QY 388 CysGlnMetThrProAlaLysPheAspGlnThrVal-----AspIleThrValGly 405  
 |||||  
 QY 1639 ACCAGTGAATTATCTGCAATCAGCCAACTG-----GCTCTGGGAATTCACATCA 1689  
 Db |||||  
 QY 406 LysGlyValPheArgValThrGlyGlnValGlnThrPheAlaGlyPheLeuSerValTyr 425  
 |||||  
 QY 1690 CTGTCATCTCAGATGACAGCCGATCAAGTGCAGGTGGCGAGATATGTCACAGACTTA 1749  
 Db |||||  
 QY 426 GluGluSerSerAspGluGlu-----GlyGluAsp----- 436  
 |||||  
 QY 1750 GATGAATGGATCTGCTGACATCTCTGCCCAATCTGAGTACCAAGCTCAGCGAATAT 1809  
 Db |||||  
 QY 437 -----SerLysLysLeuProGluMetSerGlu----- 445  
 |||||  
 QY 1810 GTTCTGTCCAGATCATTTCTTGAGGATACCACTCTCTGCTCAGCTTA----- 1860  
 Db |||||  
 QY 446 -----GlyAspLysLeuProValAspLysLeuTyrGlyGlu 457  
 |||||  
 QY 1861 CAGTATATCACTAGTTCTATGACCAATGCCCCAAGGCGCAGAGCTGTAGTGTTC 1920  
 Db |||||  
 QY 458 GlnHisPheThrPro-----ProProArgTyrAsnGluAlaThrLeuVal 473  
 |||||  
 QY 1921 TTCAGTCTGGTGTCTGTAACATG-----GCCTCTCCCAACAGCTGTTCAC 1968  
 Db |||||  
 QY 474 LysAlaLeuGluGluTyrGlyIleGlyArgProSerThrTyrAlaSerIleSerThr 493  
 |||||  
 QY 1969 AAGAGCTCTCTGAGTACGAGCTCTGGACCAACATTCACAGCTGTGTTCCATAT 2028  
 Db |||||  
 QY 494 LeuLysAspArgGluTyrValThrLeuGluGlnLys----- 505  
 |||||  
 QY 2029 CTACGATCCATCTTACAGGATTTAAGCACTTGAATACTTAATTCAGAAACGGCAGT 2088  
 Db |||||  
 QY 506 -----ArgPheMetProThrAspThrGlyAsp 514  
 |||||  
 QY 2089 GTGATTGTGAATAGCAAAATG-----AAGTTTGTAACTGTGTCGGTATAACCTCACC 2142  
 Db |||||  
 QY 515 ---IleValAsnLysPheLeuThrGluHisPheAlaGlnTyrValAspTyrHisPheThr 533  
 |||||  
 QY 2143 AAGCTGTGTCAGGGTCTTGGAGGATTTTCTGTCGTCAGCCCAACCACTCCATCTG 2202  
 Db |||||  
 QY 534 AlalysLeuGluAspGlnLeuAspGluIleAlaAspGlyLysArgGlnTrpIleProLeu 553  
 |||||  
 QY 2203 -----GAAATAGACAGC 2214  
 |||||  
 QY 554 MetAspLysPheTrpLysProPheIleLysGlnValGluGluLysGluGlyIleGluArg 573  
 |||||  
 QY 2215 TACTCTCTCAACATTTGAACAGCTGATCAGCAGATCCC---TGC----- 2256  
 |||||  
 QY 574 AlalysPheThrThrGlnGlnLeuAspGluThrCysProLysCysGlyGluHisLysLeu 593  
 |||||  
 QY 2257 -----AAGTTCTGCGCTGCGGCAATTTGCCAATGTGTA 2292  
 |||||  
 QY 594 GlnIleLysPheGlyLysMetGlyArgPheValAlaCysAlaGlyTyrProGluCysSer 613  
 |||||  
 QY 2293 -----AAGAACGAACGAGCTGAGGAAGCGAGTGTGCTGCAACACGAGTAT 2340  
 |||||

Db 614 TyrThrArgAsnValAsnGluThrAlaGluGluAlaAlaGluArgIleAlaLysAlaGlu 633  
 |||||  
 QY 2341 GACAGCCAGGGAGCCTGACGGTCTGGAACACGAGCGCTCTGTGGCCCTGSCACAAAGGAA 2400  
 |||||  
 Db 634 AlaGluGlnAlaGluLeuAspGlyArgGluCysProLysCysGly---Gly-ArgLeuVa 652  
 |||||  
 QY 2401 TCCGAGGTCTCTCCAGGAAAGGGAGCTCCATGCGAGTTGCCAGATCACTCTGAAATCAA 2460  
 |||||  
 Db 652 ITrLysTyrSerArgThrGlySerLysPheIleGlyCysAlaAsnTyrProLysCysLys 672  
 |||||  
 QY 2461 GCATACAAACTAGTGTATAAAGTCCCAATATCAACAAATACACAGGTATCAGTAA 2520  
 |||||  
 Db 672 sHisValGlu---ProLeuGluLys-PropLysAspThrGlyValGlnCysProGlnCysL 691  
 |||||  
 QY 2521 AGAAATCTGAATTAAGTACCGCTAGATATGAAGAAATTAACCATCAGATTGCGAAGG 2579  
 |||||  
 Db 691 yLysGlyAsnLeuValGluArgLys-----SerArgTyrGlyLys 704  
 |||||  
 RESULT 9  
 A55535  
 version precursor - mouse  
 N;Alternate names: chondroitin sulfate proteoglycan 2; chondroitin sulfate proteoglycan  
 versican  
 N;Contains: glial hyaluronate-binding protein  
 C;Species: Mus musculus (house mouse)  
 C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C;Accession: A55535  
 R;To: K.; Shinomura, T.; Zako, M.; Ujita, M.; Kimata, K.  
 J. Biol. Chem. 270, 958-965, 1995  
 A;Title: Multiple forms of mouse PG-M, a large chondroitin sulfate proteoglycan generated  
 A;Reference number: A55535; MUID:95122551; PMID:7822336  
 A;Accession: A55535  
 A;Status: Preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-2397 <RES>  
 A;Cross-references: GB:D16263; NID:G862460; PIDN:BA03796.1; PID:G862461  
 C;Superfamily: versican; C-type lectin homology; complement factor H repeat homology; EG  
 F;1-20/Domain: signal sequence #status predicted <SIG>  
 F;21-1654/Domain: versican #status predicted <MAT>  
 F;167-244/Domain: link protein repeat homology <LNK1>  
 F;265-346/Domain: link protein repeat homology <LNK2>  
 F;2095-2126/Domain: EGF homology <EG1>  
 F;2133-2164/Domain: EGF homology <EG2>  
 F;2171-2291/Domain: C-type lectin homology <LCH>  
 F;2298-2354/Domain: complement factor H repeat homology <FHD>  
 Alignment Scores:  
 Pred. No.: 0.0505 Length: 2397  
 Score: 152.50 Matches: 163  
 Percent Similarity: 32.31% Conservative: 110  
 Best Local Similarity: 19.29% Mismatches: 321  
 Query Match: 2.60% Indels: 251  
 DB: 1 Gaps: 34  
 US-10-007-270-1 (1-3330) x A55535 (1-2397)

QY 560 TTCAGCATTTCCAGAGCACCCTGGATCTTTCACAGACAGATAAACAACAGAGAGTTTC 619  
 |||||  
 Db 480 PheLysLysSerLysGluGlyThrGluLeuLeuGluIleThrTrpLysProGluThrTyr 499  
 |||||  
 QY 620 CTTACAGAAAAGATGAATATCTGCAGAGACAGATTCGGAGAGCTGGTGAACACCAT 679  
 |||||  
 Db 500 ProGluThrProAspHisValSerSer-----GlyGluProAspValPhePro 515  
 |||||  
 QY 680 GTCATTTCAACAGCAATCTCATTTCAAGACTTGGCGCAGTATTTCAAGAAAACCTCAG 739  
 |||||  
 Db 516 ThrLeuSerSerHisAspGlyLysThrArgTrp--SerGluSerIleThrGluSerS 535  
 |||||  
 QY 740 AAGAGCAATTAACAGATGTTCCACAGTCTCACTTGGCGCTTCCCTCTCACTCTCGATG 799  
 |||||  
 Db 535 erProAsnLeuGluAsnProValHisLysGlnProLysProValProLeuPheProGluG 555  
 |||||



```

Db      1223 euValThrGluSerGlyAlaalaargSerValLeuMetAspSerSerThrGlnGlu----- 1241
QY      2465 ACRAAACTAGTGTAAAGATTCAAA-----AATCAACAAATAACA 2506
Db      1242 --GluGluSerIleIysLeuPheGlnLysGlyValLysLeuThrAsnLysGluSerAsnA 1261
QY      2507 AGGTAATCAGT 2517
Db      1261 laAspLeuSer 1264

RESULT 10
G64529
outer membrane protein - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C;Accession: G64529
R;Tomb, J.F.; White, O.; Kervilave, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne-
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: G64529
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-595 <TOM>
A;Cross-references: GB:AE000529; GB:AE00511; NID:g2313152; PIDN:AAD07150.1; PID:g231316

Alignment Scores:
Pred. No.: 0.0486 Length: 595
Score: 150.50 Matches: 113
Percent Similarity: 35.90% Conservative: 64
Best Local Similarity: 22.92% Mismatches: 198
Query Match: 2.56% Indels: 118
DB: 2 Gaps: 20

US-10-007-270-1 (1-3330) x G64529 (1-595)
QY      1000 ATGCARAAAGTATTAAAGAACTTCAGGATTCAGAAATCAAAATATCCATGTGTAGATTAGA 1059
Db      1 MetLysLysSerPheLysLysLeu--GlyPheValSerLeuAlaAlaSerGlyValLeu 19
QY      1060 CCAAGAAAGAAAAGATGGCTCAAGCTCCACAGAGATGCAATTCAGGCACTTTAAAG 1119
Db      20 -----LeuGlySerMetAsnAlaThrAspLeuGluThrTyraAlaLeuGln 35
QY      1120 AGACACAGT-----GCAGAAAGCAAAAGCCCTGCAAGTGACCTCCTG 1161
Db      36 LysSerSerHisValPheGlyAsnTyraAlaGlu--LysAspLysAspSerLysLeuThr 54
QY      1162 TCCTTTGATTCACAAATTTGAACTGAGGAAGTCTATCATGGAACCATCGAGGAGGAC 1221
Db      55 SerAspSerProThrGlnGlnAspGlnLysValAlaGlnAsnThrAlaSerAsnAsp 74
QY      1222 AAGCAACCAAGAAATCTATCTCACAGCTACAGACCTCAAGAGCTCATCGAAGCACTA 1281
Db      75 SerGlnGlu-----AlaThrThrLeu 81
QY      1282 GAGGAAGAACAACTTTTGGATGTGGGAGCAATTCAGTTCACTGATGAA----- 1329
Db      82 GluAsnThrAlaSerThrAsp--AsnThrThrAlaThrAspGluThrTyThrLys 100
QY      1330 -----ATTGCTGATCACTGCCAGCCTTTGTCCTGACCAACCCCAATCAGAG 1374
Db      101 SerThrAspThrThrValAlaGlyAlaAlaGlnLysValGluThrAsp----- 116
QY      1375 CTGCCCAACATCTTTTGTGTATAACAGAGAGTCTACTTTGATGTCAGAACTTCCTCT 1434
Db      117 -----AsnThrAlaValGlnSerAlaGluGlnThrLysThrAspValAlaLys 133
QY      1435 GTTGAACCCAGCTTGAGACAGTGACGAGGAGCAGACATGGTCTACCTGACACTCTTGG 1494

```

RESULT 11

I52257

episialin - mouse

C;Species: Mus musculus (house mouse)

C;Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 05-Nov-1999

C;Accession: I52257; I65210

R;Vos, H.L.; De Vries, Y.; Hilken, J.





QY 2242 CAAGCAGATCCC-----TGCAAGTTC 2262  
 Db |||: |||  
 QY 527 GlnSerArgProGlyValProGlyTrpGlyIleAlaLeuValLeuValCysIleLeu 546  
 Db |||: |||  
 QY 2263 CTGGCTGGCGC-----GAATTTGCCCAATGTCTAAAGACGACGAGCT 2307  
 Db |||: |||  
 QY 547 ValAlaLeuAlaIleValTyrPheLeuAlaLeuAlaValCys----- 560  
 QY 2308 GAGCAAGCGAGTCTGCTGCAACACAGGATATGACGACGAGGAGCTGGAC 2361  
 Db |||: |||  
 QY 561 -----GlnCysArgArgLys-----SerTyrGlyGlnLeuAsp 571  
 Db |||: |||  
 RESULT 12  
 T26517  
 hypothetical protein Y18D10A.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T26517  
 R:Harris, B.  
 submitted to the EMBL Data Library, December 1998  
 A:Reference number: Z20226  
 A:Accession: T26517  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1634 <WIL>  
 A:Cross-references: EMBL:AL034393; PIDN:CAA22308.1; CESP:Y18D10A.1  
 A:Experimental source: clone Y18D10A  
 C:Genetics:  
 A:Gene: CESP:Y18D10A.1  
 A:Introns: 7/3; 44/3; 106/3; 193/1; 500/3; 533/3; 560/3; 590/3; 709/3; 1218/3; 1318/2; 1

## Alignment Scores:

Pred. No.: 0.0843 Length: 1634  
 Score: 148.50 Matches: 181  
 Percent Similarity: 33.11% Conservative: 118  
 Best Local Similarity: 20.04% Mismatches: 356  
 Query Match: 2.53% Indels: 249  
 DB: 2 Gaps: 35

US-10-007-270-1 (1-3330) x T26517 (1-1634)

QY 504 ACTGGTTCAGATCTGCCAGCAGACCTTCTGCTCTTTGATGTAAGAAACTTCA 563  
 Db |||: |||  
 QY 583 ThrMetSerIleThrGluAspSerGluProSerThrSerThrAlaAlaLys----- 600  
 Db |||: |||  
 QY 564 GCATTTCCAGCAGCAGCTGATCTTCTCCAGCAGAAATATAACAGAGAGTTTCCCTG 623  
 Db |||: |||  
 QY 601 -----ArgSerLysArgAlaGluSerAspGluGluGluGluGluAspLeuLysLeu 617  
 QY 624 ACAGAAAGATGAATATCTGCAGAGACATTTGGGAGAGCTGGTGAACCATTTGTC 683  
 Db |||: |||  
 QY 618 ThrAsnLysSerPro-----GluLysProLysLysPro----- 628  
 QY 684 TTTCACACGATTCATTTCAAGACTTTGGGAGTATTCTAAGAAACCTCAGAGA 743  
 Db |||: |||  
 QY 629 -----SerLysThrThrGluGluThrValGlyAspValLeuLysLysArgLeuArg-As 646  
 QY 744 GCAATTCAGATTTGCCAGCTCTCACTTGGGCTTTCCCTCTCACTCTGATGACAC 803  
 Db |||: |||  
 QY 646 pThrAlaLysThrThrAlaThrValIleHisThrProGlyPro-----ProLeuArgTh 664  
 QY 804 CTCTCTCAATGAATTCCTC----- 822  
 Db |||: |||  
 QY 664 rArgLysMetGluArgMetArgAlaProThrAlaValThrSerSerLysLysGluLysPr 684  
 QY 823 -----GATAATACATCAACGACACCAAGATGCTACACAGAAAG 863  
 Db |||: |||  
 QY 684 oLysAsnAlaGlySerAlaAspSerSerIleAsnGlu-----GluGluHi 699  
 QY 864 AGAAACAGAAATTCGCTGTTGGAGCAGAGGTTGGAGCTCAGCTCTCTCTGTTAA 923  
 Db |||: |||  
 QY 699 sGluAspGluThrMetIleLeuGluGluGlnThrLeuAspLeuProGlnGlnThrSerGl 719

QY 924 CCAGAAGTTTCAAG-----GCAGAGTCTGCTGACTCCAGTCCCACTATATACCA 971  
 Db |||: |||  
 QY 719 rGlnGluProArgIleSerCysGlySerGluLeuLeuAspGluGln----- 734  
 QY 972 GGAGCTAGCAGGAAAGTCCCACTTCAGATGCAGAAAGATATTTAAGAAATCCAGGATT 1031  
 Db |||: |||  
 QY 734 ----- 734  
 QY 1032 CAAAAAATCCATGTGTTAGGATTATTAGACCAAGAAAGAAAGATGGCTCAAGCTCCAC 1091  
 Db |||: |||  
 QY 735 -----PheAspAlaSerGluGluHisSerGlyThrValProSe 747  
 QY 1092 AGAGATGCAACTTACGCCCATCTTTAAGAGACACAGTGCAGACAGAAAGCCCTGCAAG 1151  
 Db |||: |||  
 QY 747 rAlaProGluLeuThr-----LysAsnProAlaPr 757  
 QY 1152 TGACCTCTCTGCTCTTTGATTCCCAAAATTTGAAGTGGAGAGTCTATCATGGAACCAT 1211  
 Db |||: |||  
 QY 757 oProValProGluAlaSerGluAlaSerAlaGluProProLysIle-----AspI1 774  
 QY 1212 GGAGAGGAGCAGCAACAGAAATCTATCTCAGAGCTCAGACCTCAAAAGGCTGATCAG 1271  
 Db |||: |||  
 QY 774 eProGluGlnAlaThrPro-----IleLeuAlaLeuAlaLeuAlaLeuProThrValSerPr 793  
 QY 1272 CAAAGCAGCTAGCAGAGCAACATCTTTGGATGGGGGACAAATTCAGTTC-----ACTGA 1325  
 Db |||: |||  
 QY 793 oThrAlaLeuGluProProLysAlaGlnGluAsnProThrAlaGluLeuProThrThrSe 813  
 QY 1326 TGAATTCCTGATCACTGCGACGCTTTGGTCTCTGACACCCCAANTCAGAGCTGCCACATC 1385  
 Db |||: |||  
 QY 813 rGluIleSerGlyArgAlaProGlnAlaLeuProThrSerSerGlnThrProProThrSe 833  
 QY 1386 TTTTCTCTTATAACAGAGAGTGTCTTCTGAGTCCAGAACTCTCTCTCTTGAACCCCA 1445  
 Db |||: |||  
 QY 833 r-----GlySerAlaAlaProProValAspLys 843  
 QY 1446 GCTT-----GAGACAGTGCAGGAGCA----- 1467  
 Db |||: |||  
 QY 843 uLeuSerGluIleLeuSerGlyAlaLysThrThrLysThrArgLysAlaAlaProAla 863  
 QY 1468 -----GAGCATGGTCTACCTGACACACTTCTGTCTCTCC 1499  
 Db |||: |||  
 QY 863 aValGlnLysSerIleSerSerThrThrGlnGlnAlaProProThrSerValGlnAlaPr 883  
 QY 1500 ACTGCTATGGCTCTTACCTCTCTGTCAGAGCTCCACCTTCTTCTTATGGCATCAAGAT 1559  
 Db |||: |||  
 QY 883 oPro-----ThrSerCysSerAlaAlaProPro-----ValAspAspLe 896  
 QY 1560 CTTCCTCTCTGATGATCAAGGACCCACAGATACATGCGCCTGACACAGCATGCTAGT 1619  
 Db |||: |||  
 QY 896 uLeuSerGluIleLeuSerGlyAlaLysThrThrLysThrThrLysMet-- 915  
 QY 1620 ACCAGGCTCACCATCCCACTGATATTCTGCAATCAGCAACTGGCTCTGGGAAT 1679  
 Db |||: |||  
 QY 916 -----ProProValAspGlnLysIleSerSerGluAlaProProI1 930  
 QY 1680 TTCAATCCACCTGATCTTCAGATGACGCGATCAAGTGCAGTGGCGAGATATGGT 1739  
 Db |||: |||  
 QY 930 eSerAspSerAlaProThrSerValHisGlnGlnThrProLysSerProLysGlnIleLe 950  
 QY 1740 CAGACACCTAGATGAATGATCTGTCGACACTCTCGCCCATCTGAGGTACAGAGCT 1799  
 Db |||: |||  
 QY 950 uAsnSerLysTyrGlyLeuAspIleSerAspSerGluAspGluGluGluGluGlu 970  
 QY 1800 CAGCAATATGTTCTCTCCAGATCATTTCTTGGAGATACCACTCTCTGCTCAGCTTT 1859  
 Db |||: |||  
 QY 970 uArgGlyMetGluIleValGluGluGlu-----GluGluAlaProProIleSerAspSe 988  
 QY 1860 ACAGTATATCACCAGTATGTTCTATGACCATTTGCCCAAGGCCGAGAGCTGTAGTGT 1919  
 Db |||: |||  
 QY 988 rLeuGlnAlaSerGluProSerThrAlaThrValLysProGluLysValValAlaVal 1008

QY 1920 CTTCAAGTCTGCGTGTGTTGCTACATGGCTTCTCCACAGCAGCTCTTCAACAGAGC----- 1974  
 Db 1008 lValysIle-----PheSerProGluIleAaspSerThrSerValGI 1022  
 QY 1975 -----TCTCTGGAGTACCG 1988  
 Db 1022 uAlaProProGluAlaSerValProProLysAlaProAlaIleThrLysIleIlePaspAsnGI 1042  
 QY 1989 AGCTCTGGAGCAACATTCACAGCAGCTGCTGTTTCCATATCTACGATCCATCTTACAGG 2048  
 Db 1042 lIleuAlaGspGlnAlaSerGluProLysAlaArgLysLeuProIleal 1062  
 QY 2049 ATTAAAGCACTTGAATATCTTAACTTCAGAACCGGAGTGATGTGATAGCAAAAT 2108  
 Db 1062 aArgLys-----IleProProLysIle 1069  
 QY 2109 GAATTTGCTAGTCTGCGGATTAACCTCACCAGGCTGTGACGGGGTCTTGAGGA 2168  
 Db 1069 eLysIleSerLeuProAlaProSerSerSerThrThrSerAaspAspLeuMetSerAs 1089  
 QY 2169 TTTTCTCTCTGCTCA-----GCCCAACAACCTCCATCT 2201  
 Db 1089 lIleLeuAlaGlyAlaLysThrThrLysThrThrLysProLysAlaProProThrArgVa 1109  
 QY 2202 GGAATACAGACGTACTCTCTCAATTTGAACCGAGCTGATCAGCAGATCCCTGCAAGTT 2261  
 Db 1109 lAlaGlnThrThrArgThrLysAsnLeuAlaGlnLysArgLysAlaSerPro----- 1126  
 QY 2262 CTTGCGCTGGCGGAAVTTGCCAATGTGTAAAGAACGACGAGTGTGAGGAGCGAGTG 2321  
 Db 1127 -----ProThrProAlaGlyThrTh 1133  
 QY 2322 TCGTGTCAACACGATATGACAGCGAGGAGCTGTGACGCTGTGACACGAGGCTCTG 2381  
 Db 1133 rAlaProLysArgGlnThrLysLysSerLysSerValLeuProProSerSe 1152  
 QY 2382 TGGCCCTGGCACAAG-----GAATGCGAGGCTCT 2411  
 Db 1152 rSerSerSerThrGluProProSerAlaProAaspSerAlaSerThrThrSerSerMetLy 1172  
 QY 2412 CCAGGGAAGGAGCT-----CCATGAGGTTGCCAGATCACTCTGA 2453  
 Db 1172 lLysGlyGlyAlaIleMetIleGluAlaValProLysArg--ProGlyGlyLys-- 1190  
 QY 2454 AAATCAAGCATACAAACTAGTGTAAAGTTCCAAATCAACAAATPACAGGTAAT 2513  
 Db 1191 -----AlaIleLysArgGlnLysProIleGlyMetLysGluMe 1204  
 QY 2514 CAGTAAAGAAATCTGAATT-ACAGCGGTAGAAATATGAAGAAATTAACCATCAAGATT 2572  
 Db 1204 tMetValGlnAsnValGluLysGlyLysLysVal----- 1216  
 QY 2573 GGGAAAGAAATTAACAACTGAAATGTACAA--TTATCTTAGCTATCTCAAGAGAG 2629  
 Db 1217 -----AsnLysIleLysThrHisLeuArgGlnAlaLeuAspLeuLysIleProPheGluGI 1235  
 QY 2630 ATGATTGCTCTTCAAGGAAATG--CAGACAGCATATTCATGGTCTCATCAATTC 2686  
 Db 1235 u-----LeuLysArgProMetGluLysGlyLysLeuLysLysSerIlePr 1252  
 QY 2687 AGCATACAGTCA--ACACTGAGAAATCAGACACACCATATTTCAAAATATAGAAGATC 2743  
 Db 1252 oLeuThrProSerAspAlaValaAspValMetGluPheLeuArgGluThrSerAlaI 1272  
 QY 2744 ATGTACTTGGCAACAGTAAATTCGAAATAAAGACACACTTACTTATTATTAACCCCA 2803  
 Db 1272 aAspMetTrpAlaValLeuAsnArgGlnArgIleAspAlaasn-----LeuLysProLe 1290  
 QY 2804 AATGCAATCAGCGAAACATATTTTACTTCTTGGATGATAGTCAAAATCATCATAGC 2863  
 Db 1290 uMetAsnLysGluGluAsnPhe-LeuGlnValSerValThrLeuAsnGluHisAspG 1310  
 QY 2864 CAGTTTGTCTCCACCTTCCCTGAAATATTTTACTCAGATCATTTTGCACAAGCATAGC 2923

Db 1310 lGlnLeuLeuGluValPheIleSerArgIleIlePheGlu-LeuCysAlaArgGluSer 1329  
 QY 2924 TTA 2926  
 Db 1330 Ile 1330  
 RESULT 13  
 S74668  
 hypothetical protein sll1665 - Synecocystis sp. (strain PCC 6803)  
 C:Species: Synecocystis sp.  
 A:Variety: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
 C:Accession: S74668  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yanada, M.; Yasuda DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis s.  
 A:Reference number: S74322; MUID:97061201; PMID:8905231  
 A:Accession: S74668  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-589 <KAN>  
 A:Cross-references: EMBL:D90900; GB:AB001339; NID:9151768; PIDN:BAA16820.1; PID:d101755  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 Alignment Scores:  
 Pred. No.: 0.0709 Length: 589  
 Score: 148.00 Matches: 89  
 Percent Similarity: 36.87% Conservative: 57  
 Best Local Similarity: 22.47% Mismatches: 136  
 Query Match: 2.52% Indels: 114  
 DB: 2 Gaps: 19  
 US-10-007-270-1 (1-3330) x S74668 (1-589)  
 QY 802 ACCCTCTCTCAATGAATTTCTCGATATATACA-----CTCAACGACACCAAGATGCT 852  
 Db 27 ThrLysLeuGlnLysLysLeuAspSerThrLysSerLysLeuGlyArgAlaLysSerAla 46  
 QY 853 ACACAGAAAGAGAGAAACAGAAATTCGCTGCTGGAGGAGCAGAGGGTGGAGCTCAGCGTC 912  
 Db 47 SerGluAspThrGluThrLeuGlnPheGlnLeuGluAlaGlnLysAlaGlu----- 63  
 QY 913 TCTCTGTAAACAGAGAGTTCACAGAGCTCGCTGACTCCAGTCCCATATTTACCAG 972  
 Db 64 -----ThrGlnAlaLeuGlnAlaArgLeuGluSerLeuGluSerSerHisArgAsn 80  
 QY 973 GAGCTA-----GCAGAAAGTCCCACTT----- 996  
 Db 81 GlnLeuAspGluLeuGlnThrAlaGlnAlaAlaAlaLeuAlaGlnLeuThrAlaGluLys 100  
 QY 997 -----CAGATCCAAAAGATATTTAAGAAACTTCCAGATTCAGAAATATC 1041  
 Db 101 ThrAlaLeuGluAspGluLeuGlnSerLeuArgGluSerGlyProAlaGluGln----- 119  
 QY 1042 CATGTGTAGGATTTAGACCAAGAGAAAGAAAGATGGCTCAAGCTCCACAGATGCA 1101  
 Db 120 -----ArgProLeuAspGlnSerAspGluValAlaAlaLeuGlnAlaGln 134  
 QY 1102 CTTACGCCCATCTTTAAGACACACAGTGCAGAGCAAGCAAAAGCCCTGCAAGTGCCTCTG 1161  
 Db 135 LeuLeuAlaMetGluGlnGluHisGlnAlaArgValGlnAla-----LeuGln 150  
 QY 1162 TCTTTGATTCACAAAATTTGAAGTGGAGGAGTCTATCATGGAACCATCAGAGGAGAC 1221  
 Db 151 AsnGlnTyxGlnAlaGluIleGluSerLeuGlnGlnAlaGlnProSerValGluGluGln 170  
 QY 1222 AAGCAACCAAGAAATCTATCTCACAGCTACAGACCTCAAAAGGCTGATCAGCAAGACACTA 1281  
 Db 171 GluGluPro-----Leu 174

QY 1282 GAGGAGAGCAACATCTTGGATGTG-----GGGCAATTCAGTTCACATCAT 1326  
 Db |||||:||||| ||| :||| |||||  
 QY 1327 GAAATCTCGATGATCTCCAGCCCTTGTCTGACACCAATCAGAGTCCACATCT 1386  
 Db |||||:||||| ||| :||| |||||  
 QY 1387 TTTCTCTTATACAGAGATGCTACTTTCAGTCCAGCAATCTCTCTGTTGAAACCCAG 1446  
 Db |||||:||||| ||| :||| |||||  
 QY 1447 CTGACAGAGTGGAGGAGAGAGATGCTTACTGACACTCTCTGG-----TCTCCA 1500  
 Db |||||:||||| ||| :||| |||||  
 QY 1501 CTTGCTATGGCTCTACTCTCCCTGTCAGAGCTCCACTTCTTTATGGCATCAAGATC 1560  
 Db |||||:||||| ||| :||| |||||  
 QY 1561 TTCTCTCTGATCAAGGACCAACAGATCAATGCGCATCTGACAGACAATGCTAGTA 1620  
 Db |||||:||||| ||| :||| |||||  
 QY 1621 CAGGGCTCCATCCCTCC-----ACCAGTGATTTCTGCAATCAGCCACTGGCTCTG 1674  
 Db |||||:||||| ||| :||| |||||  
 QY 1675 GGAATTCACATCCACTGATCTTCAGATGACAGCCGATCAAGTGGAGTGGCGAGAT 1734  
 Db |||||:||||| ||| :||| |||||  
 QY 1735 ATGGTCACACACTAGATGAATGATCTCTGACACTCTCTGCGCCCATCT----- 1785  
 Db |||||:||||| ||| :||| |||||  
 QY 1786 -----AsnPheGluProMetAspLeuAala-----ThrGluLeuProAspIleAsnAsp 320  
 Db |||||:||||| ||| :||| |||||  
 QY 321 GluGlyIleThrAsnProGlyGlnLeuAlaGluLeuSerValGluThrProGlu 340  
 Db |||||:||||| ||| :||| |||||  
 QY 1798 CTCAGCAATATGTTCTGCTCCAGATCAATTTCTGGAGGATACCACT 1845  
 Db |||||:||||| ||| :||| |||||  
 QY 341 AlaAlaGluThrGluGlnSerSerAspGlnPheLeuAlaGluLeuThr 356  
 Db |||||:||||| ||| :||| |||||  
 RESULT 14  
 S48478  
 glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)  
 N;Alternate names: extracellular glucoamylase; mucin-like protein MUC1; protein YIR019c  
 C;Species: Saccharomyces cerevisiae  
 C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text\_change 12-Nov-1999  
 C;Accession: S48478; A26877; B26877; S27281; JC6123  
 R;Rowley, K.  
 submitted to the EMBL Data Library, October 1994  
 A;Reference number: S48478  
 A;Accession: S48478  
 A;Molecule type: DNA  
 A;Residues: 1-1367 <ROW>  
 A;Cross-references: GB:247047; EMBL:M16164; NID:9603997; PID:g763364; GSPDB:GN00009; MIF  
 J. Yamashita, I.; Nakamura, M.; Fukui, S.  
 J. Bacteriol. 169, 2142-2149, 1987  
 A;Title: Gene fusion is a possible mechanism underlying the evolution of STAL.  
 A;Reference number: A91831; MUID:87194600; PMID:3106330  
 A;Accession: A26877  
 A;Molecule type: DNA  
 A;Residues: 1-242 <YAM>  
 A;Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525  
 A;Accession: B26877  
 A;Molecule type: DNA  
 A;Residues: 762-1331 <YA2>  
 A;Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526  
 R;Pardo, J.M.; Ianez, B.; Zalacain, M.; Clarios, M.G.; Jimenez, A.  
 FEBS Lett. 239, 179-184, 1988  
 A;Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar

A;Reference number: S27281; MUID:89031230; PMID:3141213  
 A;Accession: S27281  
 A;Molecule type: DNA  
 A;Residues: 1-31 <PAR>  
 A;Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552  
 R;Lambrecht, M.G.; Bauer, P.F.; Marmur, J.; Pretorius, I.S.  
 Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996  
 A;Title: Muc1, a mucin-like protein that is regulated by Msi10, is critical for pseudohy  
 A;Reference number: JC6123; MUID:96323237; PMID:8710886  
 A;Accession: JC6123  
 A;Status: nucleic acid sequence not shown  
 A;Molecule type: DNA  
 A;Residues: 1-1367 <LAM>  
 A;Cross-references: GB:U30626; NID:g1304386; PIDN:AA49609.1; PID:g1304387  
 C;Genetics: SGD:MUC1; STA2; MAL5; DEK2; SGD:S0001458  
 A;Gene: SGD:MUC1; STA2; MAL5; DEK2; SGD:S0001458  
 A;Cross-references: MIPS:YIR019c; SGD:S0001458  
 A;Map position: 9R  
 C;Superfamily: Yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase  
 C;Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein  
 F;5-21/Domain: transmembrane #status predicted <TML>  
 F;1350-1366/Domain: transmembrane #status predicted <TM2>

Alignment Scores:  
 Pred. No.: 0.101 Length: 1367  
 Score: 147.00 Matches: 108  
 Percent Similarity: 36.09% Conservative: 58  
 Best Local Similarity: 23.48% Mismatches: 194  
 Query Match: 2.50% Indels: 100  
 DB: 1 Gaps: 18

US-10-007-270-1 (1-3330) x S48478 (1-1367)

QY 830 CACTCAACGACCAAGATGCTTACACAGAAAGAGAGAAAGATTCGTGTGTGGAG 889  
 Db |||||:||||| ||| :||| |||||  
 QY 290 HisAspThrThrPro---CysThrLysLysThrThrSerLysThrCysThrLys 308  
 Db |||||:||||| ||| :||| |||||  
 QY 889 ----- 889  
 Db 309 LysThrThrThrProValProThrProSerSerThrThrGluSerSerSerAlaPro 328  
 QY 890 -----AGCAGAGGGTGGAGTCTCAGCTCTCTCTGTTAAACAGAGAGTCA 934  
 Db |||||:||||| ||| :||| |||||  
 QY 329 ValProThrProSerSerThrThrGluSerSerSerAlaProValThrSerThr 348  
 QY 935 AGCAGAGCTCGTGACTCCAGTCCCATATTTACAGAGCTAGCAGAGAGTCCACAC 994  
 Db |||||:||||| ||| :||| |||||  
 QY 349 ThrGluSerSerSerAlaProValPro-----ThrProSer-SerSerThrThrGluSe 366  
 QY 995 TTCAGATGCAAAAGATTTTAAGAACTTCCAGGATTCAAAATTCATGTGTTAGGAT 1054  
 Db |||||:||||| ||| :||| |||||  
 QY 366 rSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaPro-ValThr 384  
 QY 1055 TTAGACCAAGAAAGAAAGATGGCTCAAGCTCCACAGAGATCCAACTTACGGCATCT 1114  
 Db |||||:||||| ||| :||| |||||  
 QY 385 -----SerSerThrThrGluSerSerSerAlaProValP 396  
 QY 1115 TTAAGACACAGCTGACAGCAAAAGCCCTGCAAGTGCACCTCTCTTTGATTCGA 1174  
 Db |||||:||||| ||| :||| |||||  
 QY 396 roThrProSerSerSerThrThrGluSerSerSerAlaProValThrSer---SerThr 415  
 QY 1175 ACAAAATGAAAGTCCAGCACTCATCATGACCATGAGAGAGAGAGAGAGAGAGAG 1234  
 Db |||||:||||| ||| :||| |||||  
 QY 415 hrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaProVal 435  
 QY 1235 TCTATCTCAGCTACAGACCTCAAAAGGCTGATCAACAAAGCACTAGAGAGAGAGAG 1294  
 Db |||||:||||| ||| :||| |||||  
 QY 435 hrSerSerThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSer 455  
 QY 1295 CTTTGATGTGGGAGCAATTCAGTTCCTGATGAATTCGTGGATCACTGCCAGCCCTTG 1354  
 Db |||||:||||| ||| :||| |||||  
 QY 455 eAlaProValProThrProSerSerSerThrThrGluSerSerSerAlaProValThrS 475

```

QY 1355 GTCCTGACACCCAAATCAGAG-----CTGCCACA----- 1383
Db 475 exSerThrThrGluSerSerSerAlaProValProThrProSerSerSerThrThrGlu 495
QY 1384 --TCTTTTGTCTTAAACAGAGGATCTACTTTGAGTCCAGAACTTCTCTGTTGAAC 1441
Db 495 exSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaProValProThrP 515
QY 1442 CCCAGCTTGACAGCAGTGGACGAGCAGATGGTCTACCT-----G 1483
Db 515 roSerSerSerThrThrGluSerSerSerAlaProAlaProThrProSerSerSerThrT 535
QY 1484 ACATCTTGGTCTCCACCTGCTATGCTCTTACCTCTTACCTCTCCAGAGCTCCACCTTCT 1543
Db 535 hrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaProValP 555
QY 1544 TTATGGCATTACAGCATCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1600
Db 555 roThrProSerSer-----SerThrThrGluSerSerSerThrProValThrSerSerT 573
QY 1601 CTGACACGACAAATGCTAGTACAGGGCTCACCATCCCAACCCAGTATTTCTGCAATCA 1660
Db 573 hrThrGluSerSerSerAlaPro-----ValProThrProSerSerSerThrThrG 590
QY 1661 GCCAATGCTCTGGGAATTCACATCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1720
Db 590 luSerSerSerAlaProValProThrProSerSerSerThrThrGluSer----- 606
QY 1721 CAGGTGGGAGATATGCTGACACCTAGTAAATGATCTGCTGCTGCTGCTGCTGCTGCTG 1780
Db 607 -----SerSerAlaProAlaP 612
QY 1781 CATCTGAGTACAGAGCTCAGCAATATGTTCTGCTGCCAGCATCTTCTGGAGGATA 1840
Db 612 roThrProSerSerThrThrGluSerSerSerAlaPro-----ValThrSerSerT 630
QY 1841 CCAGT-----CCTGCTCAGCTTTACAGTATATCACCACCT-----AGTT 1879
Db 630 hrThrGluSerSerSerAlaProValProThrProSerSerSerThrThrGluSerSerS 650
QY 1880 CTATGACCATTTGCC-----CCAGGGCCGAGAGCTGTTGTTCTTCACTGCTGCTGTTG 1936
Db 650 exAla-ProValProThrProSerSerSerThrThrGluSerSerSerAlaProVal--- 668
QY 1937 CTAACTGGCTTCTCCACGACCTGTTCAACAGAGCTCTCTGAGTACCGAGCTCTGG 1996
Db 669 -----ProThrProSerSerThrThrGluSerSerSerAlaProValThrSe 685
QY 1997 AGCAACATTCACAGCTGCTGCTCCATA-----TCTACGATCCATCT 2042
Db 685 rSerThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSer 702

RESULT 15
S61535
nucleotide-binding head-stalk protein 183K - Giardia lamblia
C;Species: Giardia lamblia
C;Date: 21-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 07-May-1999
C;Accession: S61535
R;Marshall, J.; Holberton, D.V.
J. Cell Sci. 108, 2683-2692, 1995
A;Title: Giardia gene predicts a 183 kDa nucleotide-binding head-stalk protein.
A;Reference number: S61535; MUID:96066038; PMID:7593309
A;Accession: S61535
A;Status: Preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1620 <MAR>
A;Cross-references: EMBL:X79815; NID:9871047; PID:9871048

Alignment Scores:
Pred. No.: 0.106 Length: 1620
Score: 147.00 Matches: 174
Percent Similarity: 34.50% Conservative: 150
Best Local Similarity: 18.53% Mismatches: 345

```

```

Query Match: 2.50% Indels: 270
DB: 2 Gaps: 35
US-10-007-270-1 (1-3330) x S61535 (1-1620)
QY 222 CTGAACCTAAGACATAGACATATCCCAAGAAATGAAACAACTGAAAGTACTGAAAAA 281
Db 313 illyyGluLysSerValThrileGluLysLeuMetGluLysLeuSerLysLeuLys 332
QY 282 TGTACAAAATGTCAACTATGAGACGATATTCGATTTCGCAAGCATCCCAAAAAGAT 341
Db 333 ThrSerGlnVal-----AlaileGlnArgAsp 342
QY 342 CGGCATTTTCCCA-----CGGGGTAAAGTCTGTCCACAG----- 380
Db 343 GlnAspAlaValGluLeuAlaArgSerValAlaAspMetArgGlnLysLeuSerAspThr 362
QY 381 ---AATCCATGAACAGATTTTAGACAGCTTCAGAGCTTATTATGATTTAGAGTGTGTC 437
Db 363 ValAspGluLysGlnArgile-----ValHisGlnLeuHisileAspMetAspSerMet 380
QY 438 AGGAGCAGTATGGGAACATATCGGATCTTTC----- 470
Db 381 LysLysSerHisGluSerLysSerGlnHisLysSerGluLysSerGluLysLysSer 400
QY 470 ----- 470
Db 401 AlaPheAspGluSerThrPheileLysAsnAlaArgileSerLysLeuThrThrGluLeu 420
QY 471 -----TGGATCGCATCCCTGACACGCGGGAATATCAGAGCTGGGTGCGCA 515
Db 421 ThrGluThrGlnThrGlnLeuAlaSerAlaLeuGluLysAsnGlnGluLysGlnSerGlu 440
QY 516 TCTGCCAGCAGAGAGCTTCTGCTCTTGTGACATTTGCAAAAAAATTCAGCAATTTCCAGG 575
Db 441 SerLysLysAlaGlnHisAlaGluValLeuLeuAlaGluArgLys-----ileAlaAsn 459
QY 576 AGCAGCTGGATCTTCTCCAGCAGAGAAATAAACAAGAGAGTTTCCTGACAGAAAGATG 635
Db 460 ThrThrLeuThrGlnGluLeuGlnAlaThrLysGlyGluLeuGluAlaValArgLysSer 479
QY 636 AATATCTGACAGAGAGACATTTGGAGAGCTGCTGCAAAACCTCAGAGAGCAATTCAGA 755
Db 480 AsnHisGluGlnValAsnSerTyGluSerLeuLeuGln-----LysThrGln 495
QY 696 TCTACATTTCAAGACTTGGGAGCTGCTGCAAAACCTCAGAGAGCAATTCAGA 755
Db 496 GlnSerTyGluAlaLeuLeuArgLysGluAspThrileLeuLysLeuAsnLeuGlu 515
QY 756 TGTTCGCAACGCTCTCACTTGGGCTTTCCTCTCACTCTCTGATGACACCTCTCAATGA 815
Db 516 ---CysAspAsnAlaLysLysAlaPheAspHisSer----- 527
QY 816 AATCTCGATATACACTCAACGACACCAAGATGCTTACACAGAGAGAGAGAGATTT 875
Db 528 -----GluGlnSerThrAlaHisValGlnAlaLeuArgGluGluValGluArgVal 544
QY 876 CGCTGCTTGGAGGAGCAGAGGCTGAGCTCAGCTCTCTCTGTTAAAC----- 924
Db 545 ArgHisThr-SerGluLeuLeuLeuGlnGluLysMetGlyValileAsnArgLeuThrAl 564
QY 925 -----CAGAAAGTTCAAG----- 936
Db 564 aGluileGlnAlaLysLysLeuAspGlyLysArgAlaLeuAspSerLysAspThrGlnil 584
QY 937 -GAGAGCTGCTGCTGCTCCAGTCCCATATTTACAGAGAGT-----GCAGAGAA 986
Db 584 eAlaGluLeuGlnLysAsnValSerSerLeuMetGluAspLeuThrLysThrSerSerAs 604
QY 987 GTCCCAACTTCAGATGCAAAAGATTTAAAGAACTTCAGAGATTCAGAAATTCATCT 1046
Db 604 nSerSerAlaGluLeuSerArgLeuThrAsnLeuLeuGluAlaThrArgLysAspTyVa 624

```

```
QY 1047 GTTAGGATTAGACCAAGAAAGAAAGATGGCTCAAGCTCCACAGATG----- 1098
Db 624 lGlnArgLeuGluSerLysAspGluGlnLeuArglAlaThrAsnGlnTyrLysGluAs 644
QY 1099 -----CAACTTACGGCCATCTTTAAGACACACAGTCAGAGCAAAAGCCCTGCAAG 1151
Db 644 mMetGluLysLeuAspAlaValPheAlaLysLysAspAlaGluLeuSerLysGlnThrAl 664
QY 1152 TGACCTC-----CTGCTCTTTGATTCCACAAATTTGAAGTGAAGTCTATCATGG 1205
Db 664 aAlaLeuHisAlaLeuArgAspGluLeuSerGlnValLysSerAsnHisLeu----- 681
QY 1206 AACCATGGAGGAGCAAGCAACCAAAATCTATCTCACAGCTACAGACTCAAAAGCT 1265
Db 682 -----LeuGluValAspArg 687
QY 1266 GATCACCAAGCACTAGAGAGAAACAATCTTTGGATGTGGGGCAATTCAGTTCACTCA 1325
Db 687 uLeuLysGluValAsnSerAlaGlnGlnAsnThrSerMetThrLeuLeuSerAsnGlnAs 707
QY 1326 TGAATTTGCTGGATCACTGCCAGCCTTTGGTCCTGACCCCAATCAGAGTCCCAATC 1385
Db 707 pGluIle-----SerArgLeuGlyArgG 715
QY 1386 TTTTCTCTTATAACAGAGATGCTACTTTG----- 1416
Db 715 uLeuAspValValLysAsnSerAlaSerIleAsnGluValGluAlaAsnLysIleG 735
QY 1417 -AGTCCAGACTCTCTCTGTTGAACCCAGCTTGAGACAGTGGACGAGCAGATGG 1475
Db 735 nSerLeuAsnLeuLysIleSerLeuGlnGlnLeuGlnAspSerLysAsnGluAl 755
QY 1476 T-----CTACTGTACAC 1487
Db 755 aAspAlaAlaLeuLysGluLysValArgGluLeuAspThrLeuArgLeuLeuMetAspAs 775
QY 1488 TTCTTGCTCTCCACCTGTATGGCTCTACCTCCCTGTCAGAACTCCACCTTTCTTAT 1547
Db 775 pSerThrSerThrAlaAlaLysAlaAlaGlnAsnLeu----- 787
QY 1548 GGCATCAAGCATCTCTCTGACTGATCAGGACCAAGCAATGCGCACTGACCA 1607
Db 788 -----GlnAsnThrValAspLysLeuGlnGlnGlu-- 797
QY 1608 GACAATGCTAGTACCGGCTCCACATCCCAACAGTGATTTCTGCAATC----- 1659
Db 798 -----LeuSerSerValSerSerAspLysLeuAlaIleSerAlaG 811
QY 1659 ----- 1659
Db 811 uMetAsnArgValValSerGluLeuLysHisGluLeuSerThrGluLysGlnMetArgG 831
QY 1660 -----AGCCAACTGGCTCTGGGAATTCATCCACCTGCATCTTCAGATGACAG 1709
Db 831 uAlaGluAenSerArgAlaGlnLeuGlnIleSerHisLeuGluAlaGlnValLysAspAl 851
QY 1710 CCGATCAAGTCAGGTGGCGAA-----GATATGCTCAGACACCTAGA 1751
Db 851 aAlaLysAlaLysAspSerGluValThrArgLeuLeuSerAspLeuLysThrValLysG 871
QY 1752 TGAATGGATCTGCTCAGCTCTGCGCCCATCTGAGGTACCAGAGCTCAGCAATATGT 1811
Db 871 uGluLeuAlaIleValValAspGlnLysAspAlaLysIleAlaGluIleSerArgLysLe 891
QY 1812 TTCTGTCCTCAGATCATCTTTGAG-----GATACACTCTCTGCTCAGCTTACAGTATAT 1868
Db 891 uGluAspThrPheHisLysLeuGlnArgSerGluGlnThrValGluValLeuGlnAlaAl 911
QY 1869 CACCACTAGTCTATGACCAATTCGCCCAAGGGCGCA-----GAGCTGCTAGTCTTTCAG 1925
Db 911 aLysGluLysGluLeuSerValAlaLysLeuAsnThrAspGlnThrIleAlaLeuLeuAs 931
QY 1926 TCTGCGTCTGCTAACATGCGCTTC-----TCCACGACCT 1961
```

```
Db 931 nAspArgValAlaAsnLeuAlaAlaGluLeuAsnLysGlnLysAsnGluThrGluGluDe 951
QY 1962 GTTCAACACAGAGCTCTCTGGAGTACCGAGCTCTGAGCAACAATTCACACAGCTCTGGT 2021
Db 951 uLeuAlaPheLysAspLeuGlnTyrLysGlnLeuGlnLysGlnLeuGluAspSerArgTh 971
QY 2022 TCCATATCTACGATCCAACTCTTACAGATTTAAG-----CAACTTGAAT 2066
Db 971 rGluValThrGluAlaThrSerSerGlyArgAlaGluIleSerArgLeuGlnLeuGlnI 991
QY 2067 ACTTTAACTTCAGAAACGGAGCTGATCTGTGAATGAGCAAAATGAAGTTGTAAAGTCTGT 2126
Db 991 eAspAsnLeu-----GlyGluAlaLeuGlnAlaGlnIleGluTyrAlaGlnLys-- 1008
QY 2127 GCCGTATACTCACCAAGCTGTGCACCGGCTCTGAGGATTTTCGTTCTGCTGCAGC 2186
Db 1009 -----GluAspAlaLeuLysAspAspLeuAsnSerAlaLysAl 1021
QY 2187 CCAACAACCTC-----CATCTGGAATAGACAGCTACTCTCTCAACATTGAACCACTGA 2240
Db 1021 aValLeuLeuAlaGluSerAlaGluLysAsp-AlaIleIleSerThrLeuLysLysAsp 1041
QY 2241 TCAAGCAGATCCCTGCAAGTTCCTGCTGCGCGAATTTGCCAATGTGTAAGAACGA 2300
Db 1041 euThrAsnLeuArgAlaGlu--LeuLeuSerSerGluGluAlaLysAspValThrIleAl 1060
QY 2301 ACGGACTGAGGACGGAGGTGCTGCTGCACACAGGATATGACAGCCAGGAGCGCTGGA 2360
Db 1060 aArgTyrLysGln---AspCys-----GI 1067
QY 2361 CGGTCTGCAACCAAGGCTCTGTGGCCCTGGCCTGCAACAAGGAATGCGAG----- 2406
Db 1067 uAsnLeuGlnThrSerLeu-----ThrLysSerIleGluLysLysGluAl 1083
QY 2407 -----GTCTCTCAGGAAAGGAGCTCCATGCAAGTTGCCAGTCACTCTGAAATCA 2459
Db 1083 aTyrAsnIleLeuLysGlnGluPheAlaGlyTyrLysLysAspValSerAlaAlaLysG 1103
QY 2460 AGCATACAAACTAGTGTGTTAAAGTTCCAAAATCAACAAAATAACAAAGTAAATCAGTAA 2519
Db 1103 nAlaTyrGluAlaGlnIleAlaSerLeu-----ThrGlyAspLeuAlaAlaLys 1120
QY 2520 AAGAAATTCGAATTAATCTGACCGCTAGATATGAA-----GAATTAACCAT 2565
Db 1120 sLysLysSerGluGlnLeuGluMetGluIleGluArgGluMetLysHis 1136
```

Search completed: March 4, 2004, 18:55:01  
Job time : 142.5 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 4, 2004, 18:37:36 ; Search time 32.5 Seconds  
(without alignments)  
10670.376 Million Cell updates/sec

Title: US-10-007-270-1

Perfect score: 5875

Sequence: 1 taaaccaagaaggttatcct.....tactatatgacataatcaat 3330

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-O=/cgn2\_1/USPTO\_model\_n2p.model -DEV=xlp  
-DB=SwissProt\_42 -OFT=fastan -SUFFIX=rs -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-NO MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -XGAPOP=6  
-XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID        | Description                          |
|------------|-------|-------------|--------|--------------|--------------------------------------|
| 1          | 189.5 | 3.2         | 1255   | 1 MUC1 HUMAN | P15941 h mucin 1 p                   |
| 2          | 187.5 | 3.2         | 475    | 1 MUC1 HVILA | Q29435 Hylobates l                   |
| 3          | 153   | 2.6         | 3678   | 1 DMD_MOUSE  | P11531 mus musculus                  |
| 4          | 152.5 | 2.6         | 3358   | 1 PGCY_MOUSE | Q62059 mus musculus                  |
| 5          | 152.5 | 2.6         | 8797   | 1 SNEI_HUMAN | Q8rf91 homo sapien                   |
| 6          | 151   | 2.6         | 1341   | 1 ACIN_HUMAN | Q9ukv3 homo sapien                   |
| 7          | 150   | 2.6         | 1163   | 1 RTN4 RAT   | Q9jkl1 rattus norv                   |
| 8          | 149   | 2.5         | 7389   | 1 BPAL_MOUSE | Q9izt6 mus musculus                  |
| 9          | 148   | 2.5         | 1490   | 1 CRK7_HUMAN | Q9nyv4 homo sapien                   |
| 10         | 148   | 2.5         | 2738   | 1 PGCY_RAT   | Q9erB4 rattus norv                   |
| 11         | 147   | 2.5         | 1367   | 1 AMVH_YEAST | P08640 saccharomyc                   |
| 12         | 146.5 | 2.5         | 3396   | 1 PGCY_HUMAN | P13611 homo sapien                   |
| 13         | 145.5 | 2.5         | 928    | 1 NIBA_HUMAN | RC MEDLINE=90368715; PubMed=1697589; |
| 14         | 145.5 | 2.5         | 1024   | 1 POPC_BALSO | Q9rbz8 raietonia s                   |
| 15         | 145   | 2.5         | 630    | 1 MUC1_MOUSE | Q02496 mus musculus                  |
| 16         | 145   | 2.5         | 1036   | 1 P200_MYCN  | P75211 mycoplasma                    |
| 17         | 145   | 2.5         | 1957   | 1 SPOF_SCHPO | Q10411 schizosacch                   |
| 18         | 145   | 2.5         | 2869   | 1 RBP1_PLAVB | Q00798 plasmodium                    |

|    |       |     |      |              |                    |
|----|-------|-----|------|--------------|--------------------|
| 19 | 144   | 2.5 | 1468 | 1 N153_RAT   | P49791 rattus norv |
| 20 | 143.5 | 2.4 | 1062 | 1 YAI3_HUMAN | Q9y216 homo sapien |
| 21 | 143.5 | 2.4 | 1306 | 1 MSB2_YEAST | P32334 saccharomyc |
| 22 | 143.5 | 2.4 | 3680 | 1 DMD_CANFA  | Q97592 canis famil |
| 23 | 143.5 | 2.4 | 3685 | 1 DMD_HUMAN  | P11532 homo sapien |
| 24 | 143   | 2.4 | 3259 | 1 GOB1_HUMAN | Q14789 homo sapien |
| 25 | 141.5 | 2.4 | 3381 | 1 PGCY_BOVIN | P81282 bos taurus  |
| 26 | 141   | 2.4 | 2845 | 1 APC_MOUSE  | Q61315 mus musculu |
| 27 | 140.5 | 2.4 | 529  | 1 HSF1_HUMAN | Q00613 homo sapien |
| 28 | 140.5 | 2.4 | 1772 | 1 MSP1_PLAYO | P13828 plasmodium  |
| 29 | 139.5 | 2.4 | 1928 | 1 MYSL_YEAST | P08964 saccharomyc |
| 30 | 139   | 2.4 | 2145 | 1 CYAA_PODAN | Q01513 podospora a |
| 31 | 139   | 2.4 | 5120 | 1 PCLO_CHICK | Q9pu36 gallus gall |
| 32 | 139   | 2.4 | 8445 | 1 ANCI_CAEEL | Q9n4m4 caenorhabdi |
| 33 | 138.5 | 2.4 | 2230 | 1 GOA4_HUMAN | Q13439 homo sapien |
| 34 | 138.5 | 2.4 | 4128 | 1 PRKD_MOUSE | P97313 mus musculu |
| 35 | 138   | 2.3 | 1677 | 1 VIT_ACITR  | Q90243 acipenser t |
| 36 | 137.5 | 2.3 | 1609 | 1 FIG2_YEAST | P25653 saccharomyc |
| 37 | 137.5 | 2.3 | 6885 | 1 SNE2_HUMAN | Q8wxh0 homo sapien |
| 38 | 137   | 2.3 | 1676 | 1 APSA_EMENI | Q00083 emericeella |
| 39 | 136.5 | 2.3 | 2158 | 1 MY9B_HUMAN | Q13459 homo sapien |
| 40 | 135.5 | 2.3 | 917  | 1 SMOO_HUMAN | P53814 homo sapien |
| 41 | 135.5 | 2.3 | 1142 | 1 GIN4_YEAST | Q12263 saccharomyc |
| 42 | 135.5 | 2.3 | 1222 | 1 YMH5_CAEEL | P34472 caenorhabdi |
| 43 | 135   | 2.3 | 1140 | 1 YMG6_YEAST | Q04893 saccharomyc |
| 44 | 135   | 2.3 | 1464 | 1 NCO2_HUMAN | Q15596 homo sapien |
| 45 | 135   | 2.3 | 1875 | 1 MLPI_YEAST | Q02455 saccharomyc |

#### ALIGNMENTS

RESULT 1  
MUC1\_HUMAN  
ID MUC1\_HUMAN STANDARD; PRT; 1255 AA.  
AC P15941; P13931; P15942; P17626; Q14128; Q14876; Q16437; Q16442;  
AC Q16615; Q9BXA4; Q9UE75; Q9UE76; Q9UQJ1; Q9Y4J2;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Mucin 1 precursor (MUC-1) (Polymorphic epithelial mucin) (PEM) (PEMT)  
DE (Episialin) (Tumor-associated mucin) (Carcinoma-associated mucin)  
DE (Tumor-associated epithelial membrane antigen) (EMA) (H23AG) (Peanut-  
DE reactive urinary mucin) (PUM) (Breast carcinoma-associated antigen  
DE DF3) (CD227 antigen).  
GN MUC1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Pancreas;  
RX MEDLINE=90368716; PubMed=2394722;  
RA Lan M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;  
RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";  
RL J. Biol. Chem. 265:15294-15299 (1990).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RX MEDLINE=90202794; PubMed=2318825;  
RA Lichtenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.;  
RT "Episialin, a carcinoma-associated mucin, is generated by a  
RT polymorphic gene encoding splice variants with alternative amino  
RT termini".  
RL J. Biol. Chem. 265:5573-5578 (1990).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Breast carcinoma;  
RX MEDLINE=90368715; PubMed=1697589;  
RA Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,  
RA Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;  
RT "Molecular cloning and expression of human tumor-associated  
RT polymorphic epithelial mucin".  
RL J. Biol. Chem. 265:15286-15293 (1990).

RN [4] SEQUENCE FROM N.A. (ISOFORM 1).  
 RP MEDLINE=91097524; PubMed=2268309;  
 RX Lancaster C.A., Peat N., Duhig T., Wilson D.,  
 RA Taylor-Papadimitriou J., Gendler S.J.;  
 RT "Structure and expression of the human polymorphic epithelial mucin  
 gene: an expressed VNTR unit";  
 RL Biochem. Biophys. Res. Commun. 173:1019-1029(1990).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORM 5).  
 RC TISSUE=Breast carcinoma;  
 RX MEDLINE=90276413; PubMed=2351132;  
 RA Wreschner D.H., Hareuveni M., Tsarfaty I., Smorodinsky N., Horev J.,  
 RA Zaretzky J., Kotkes P., Weiss M., Lathe R., Keydar I.;  
 RT "Human epithelial tumor antigen cDNA sequences. Differential splicing  
 may generate multiple protein forms";  
 RL Eur. J. Biochem. 189:463-473(1990).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Breast carcinoma;  
 RX MEDLINE=90276414; PubMed=2112460;  
 RA Hareuveni M., Tsarfaty I., Zaretzky J., Kotkes P., Horev J.,  
 RA Zrihan S., Weiss M., Green S., Lathe R., Keydar I., Wreschner D.H.;  
 RT "A transcribed gene, containing a variable number of tandem repeats,  
 codes for a human epithelial tumor antigen. cDNA cloning, expression  
 of the transfected gene and over-expression in breast cancer  
 tissue";  
 RL Eur. J. Biochem. 189:475-486(1990).  
 RN [7]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=91033045; PubMed=1688329;  
 RA Tsarfaty I., Hareuveni M., Horev J., Zaretzky J., Weiss M.,  
 RA Jeltsch J.M., Garnier J.M., Lathe R., Keydar I., Wreschner D.H.;  
 RT "Isolation and characterization of an expressed hypervariable gene  
 coding for a breast-cancer-associated antigen";  
 RL Gene 93:313-318(1990).  
 RN [8]  
 RP SEQUENCE FROM N.A. (ISOFORM 7).  
 RX MEDLINE=95010060; PubMed=7925397;  
 RA Zrihan-Licht S., Vos H.L., Baruch A., Elroy-Stein O., Sagiv D.,  
 RA Keydar I., Hilkens J., Wreschner D.H.;  
 RT "Characterization and molecular cloning of a novel MUC1 protein.  
 devoid of tandem repeats, expressed in human breast cancer tissue";  
 RL Eur. J. Biochem. 224:787-795(1994).  
 RN [9]  
 RP SEQUENCE FROM N.A. (ISOFORMS 6; 7 AND 8).  
 RX MEDLINE=97355747; PubMed=9212228;  
 RA Osterkamp H.M., Scheiner L., Stefanova M.C., Lloyd K.O.,  
 RA Finstad C.H.;  
 RT "Comparison of MUC-1 mucin expression in epithelial and non-epithelial  
 cancer cell lines and demonstration of a new short variant form  
 (MUC-1/2)";  
 RL Int. J. Cancer 72:87-94(1997).  
 RN [10]  
 RP SEQUENCE FROM N.A. (ISOFORM 7).  
 RA Zhang L.X., Li C.H.;  
 RT "Molecular cloning of an isoform of MUC1, MUC1/Y";  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [11]  
 RP SEQUENCE FROM N.A. (ISOFORM 9).  
 RC TISSUE=Epithelial cancer;  
 RA Zhang L.X., Li C.H., Sun L.Y., Yue W.;  
 RT "Cloning of a new potential secreted short variant form of MUC1 mucin  
 in epithelial cancer cell line";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 RN [12]  
 RP PARTIAL SEQUENCE FROM N.A.  
 RX MEDLINE=8830762; PubMed=3417635;  
 RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,  
 RA Burchell J.;  
 RT "A highly immunogenic region of a human polymorphic epithelial mucin  
 expressed by carcinomas is made up of tandem repeats";  
 RL J. Biol. Chem. 263:12820-12823(1988).  
 RN [13]  
 RP SEQUENCE OF 1-150 FROM N.A. (ISOFORM 2).  
 RX MEDLINE=9008473; PubMed=2597151;  
 RA Abe M., Siddiqui J., Kufe D.;  
 RT "Sequence analysis of the 5' region of the human DF3 breast  
 carcinoma-associated antigen gene";  
 RL Biochem. Biophys. Res. Commun. 155:644-649(1989).  
 RN [14]  
 RP SEQUENCE OF 1-109 FROM N.A. (ISOFORM 2).  
 RC TISSUE=Thyroid;  
 RX MEDLINE=96183746; PubMed=8608966;  
 RA Weiss M., Baruch A., Keydar I., Wreschner D.H.;  
 RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse  
 transcriptase polymerase chain reaction of the MUC1 gene";  
 RL Int. J. Cancer 66:55-59(1996).  
 RN [15]  
 RP SEQUENCE OF 1-89 FROM N.A.  
 RC TISSUE=Lung;  
 RX MEDLINE=96181716; PubMed=8604237;  
 RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,  
 RA Lee L.N., Luh K.T., Wu C.W.;  
 RT "Mucin mRNA expression in lung adenocarcinoma cell lines and  
 tissues";  
 RL Oncology 53:118-126(1996).  
 RN [16]  
 RP SEQUENCE OF 1-46 FROM N.A. (ISOFORMS 3 AND 4).  
 RC TISSUE=Breast carcinoma;  
 RA Buluwela L., Liu Q., Luqmani Y.A., Gomm J.J., Coombes R.C.;  
 RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.  
 RN [17]  
 RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.  
 RX MEDLINE=97460054; PubMed=9312074;  
 RA Mueller S., Goletz S., Packer N., Gooley A.A., Lawson A.M.,  
 RA Hanisch F.-G.;  
 RT "Localization of O-glycosylation sites on glycopeptide fragments from  
 lactation-associated MUC1. All putative sites within the tandem  
 repeat are glycosylation targets in vivo";  
 RL J. Biol. Chem. 272:24780-24793(1997).  
 RN [18]  
 RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.  
 RX MEDLINE=99303572; PubMed=10373415;  
 RA Mueller S., Alving K., Peter-Katalinic J., Zachara N., Gooley A.A.,  
 RA Hanisch F.-G.;  
 RT "High density O-glycosylation on tandem repeat peptide from secretory  
 MUC1 of T47D breast cancer cells";  
 RL J. Biol. Chem. 274:18165-18172(1999).  
 RN [19]  
 RP POLYMORPHISM WITHIN THE REPEAT.  
 RX MEDLINE=21359366; PubMed=11350974;  
 RA Engelmann K., Balduz S.E., Hanisch F.-G.;  
 RT "Identification and topology of variant sequences within individual  
 repeat domains of the human epithelial tumor mucin MUC1";  
 RL J. Biol. Chem. 276:27764-27769(2001).  
 RN [20]  
 RP CHARACTERIZATION OF ISOFORM Y, AND MUTAGENESIS OF ASP-1116.  
 RX MEDLINE=99211485; PubMed=10197628;  
 RA Baruch A., Hartmann M.-L., Yoeli M., Adereth Y., Greenstein S.,  
 RA Stadler Y., Skotnik Y., Zaretzky J., Smorodinsky N.I., Keydar I.,  
 RA Wreschner D.H.;  
 RT "The breast cancer-associated MUC1 gene generates both a receptor and  
 its cognate binding protein";  
 RL Cancer Res. 59:1552-1561(1999).  
 RN [21]  
 RP PARTIAL SEQUENCE, AND CHARACTERIZATION OF CLEAVAGE SITE.  
 RX MEDLINE=21240104; PubMed=11341784;  
 RA Parry S., Silverman H.S., McDermott K., Willis A., Hollingsworth M.A.,  
 RA Harris A.;  
 RT "Identification of MUC1 proteolytic cleavage sites in vivo";  
 RL Biochem. Biophys. Res. Commun. 283:715-720(2001).  
 RN [22]  
 RP CHARACTERIZATION  
 RX MEDLINE=21836452; PubMed=11847293;  
 RA Wreschner D.H., McGuckin M.A., Williams S.J., Baruch A., Yoeli M.,



|    |      |   |      |
|----|------|---|------|
| QY | 1915 | GTGTCTTC-----AGTCTGCTGTGTCTAAATGAGCGCTTCTCCACAGCACTGTTCAAC      | 1968 |
|    |      |   |      |
|    |      |   |      |
| Db | 1041 | SerPhePheLeuSerPheHisIleSerAsnLeuGlnPheAsnSerSerLeuGluAsp       | 1060 |
|    |      |   |      |
|    |      |   |      |
| QY | 1969 | AGAGCTCTCTGAGTAGTCAGGAGCTGTGGAGCAACAATTACACAGACTGCTGGTTCCTCATAT | 2028 |
|    |      |   |      |
|    |      |   |      |
| Db | 1061 | ProSerThrAspTyrTyrGlnGluLeuGlnArgAspIleSerGluMetPheLeuGluIle    | 1080 |
|    |      |   |      |
|    |      |   |      |
| QY | 2029 | CTACAGCCCAATCTTACAGGAGTTTAGCACTTGAAATACTTAACTTTCAGAAACGGGAGT    | 2088 |
|    |      |   |      |
|    |      |   |      |
| Db | 1081 | TyrIysGlnGly-----GlyPheLeuGlyLeuSerAsnIleLysPheArgProGlySer     | 1098 |
|    |      |   |      |
|    |      |   |      |
| QY | 2089 | GTGATTGTGAATAGCAAAATGAAGTTTGTCTAAGTCTGTCCGTATTAACCTCACCAGGCT    | 2148 |
|    |      |   |      |
|    |      |   |      |
| Db | 1099 | ValValValGlnLeuThrLeuAlaPheArgGluGly-----ThrIleAsn              | 1113 |
|    |      |   |      |
|    |      |   |      |
| QY | 2149 | GTGCACGGGGTC-----TTGGAGGATTTTCGTTCTGCTGCAGCCCCAACACTCCAT        | 2199 |
|    |      |   |      |
|    |      |   |      |
| Db | 1114 | ValHisAspValGluThrGlnPheAsnGlnTyrIysThrGluLeuAlaSerArgTyrAsn    | 1133 |
|    |      |   |      |
|    |      |   |      |
| QY | 2200 | CTGGAATAGACAGCTACTCTCTCAACATTTGAACCA                            | 2235 |
|    |      |   |      |
|    |      |   |      |
| Db | 1134 | LeuThrIleSerAspValSerValSerAspValPro                            | 1145 |

|   |
|---|
| RESULT 2  |
| MUC1_HYLLA  |
| ID MUC1_HYLLA STANDARD; PRT; 475 AA.  |
| AC Q29435;  |
| DT 15-JUL-1999 (Rel. 38, Created)   |
| DT 15-JUL-1999 (Rel. 38, Last sequence update)  |
| DE 28-FEB-2003 (Rel. 41, Last annotation update)  |
| DE Mucin 1 precursor (MUC-1).   |
| GN MUC1.  |
| OS Hylobates lar (Common gibbon).   |
| OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  |
| OC Mammalia; Rutheria; Primates; Catarrhini; Hylobatidae; Hylobates.  |
| OX NCBI_TaxID=9580;   |
| [1]   |
| RN SEQUENCE FROM N.A.   |
| RP MEDLINE=96351712; PubMed=8747930;  |
| RX Spicer A.P., Duig T., Chilton B.S., Gendler S.J.;  |
| RT "Analysis of mammalian MUC1 genes reveals potential functionally   |
| RL important domains.";   |
| RL Mamm. Genome 6:885-888(1995).  |
| CC -1- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN  |
| CC CYTOSKELETON (BY SIMILARITY).  |
| CC -1- SUBCELLULAR LOCATION: Type I membrane protein.   |
| CC -1- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.   |
| CC -1- SIMILARITY: Contains 1 SEA domain.   |
| CC  |
| This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). |
| CC  |
| ENBL; L41589; AAA69965.1; -;  |
| OR ENBL; L41625; AAA69918.1; -;   |
| DR ENBL; L41624; AAA69918.1; JOINED.  |
| OR InterPro; IPR000082; SEA_domain.   |
| DR Pfam; PF01390; SEA; 1  |
| OR SMART; SMO0200; SEA; 1.  |
| DR PROSITE; PS00200; SEA; 1.  |
| OR Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;  |
| GW Repeat.  |
| KT SIGNAL   |
| PT 1 23 POTENTIAL.  |
| VT CHAIN 24 475 MUCIN 1.  |
| TT DOMAIN 24 380 EXTRACELLULAR (POTENTIAL).   |
| T TRANSMEM 381 401 POTENTIAL.   |
| T DOMAIN 402 475 CYTOPLASMIC (POTENTIAL).   |
| T REPEAT 102 121  |

FT REPEAT 122 141 2.  
 FT REPEAT 142 161 3.  
 FT REPEAT 162 181 4.  
 FT DOMAIN 254 371 SEA.  
 SQ SEQUENCE 475 AA; D7A699D6D68C6622 CRC64;

## Alignment Scores:

Pred. No.: 0.000324 Length: 475  
 Score: 187.50 Matches: 85  
 Percent Similarity: 39.60% Conservative: 52  
 Best Local Similarity: 24.57% Mismatches: 124  
 Query Match: 3.19% Indels: 85  
 DB: 1 Gaps: 17

US-10-007-270-1 (1-3330) x MUC1\_HYLLA (1-475)

QY 1291 CAATCTTGATGTGGGACAAATTCAGTTCATGATGAATTCGTGATCATCGCCAGCC 1350  
 Db 74 GlnGlyGlnAspValSerLeuAlaProAlaThrGluProAlaSerGlySerAlaAlaThr 93  
 QY 1351 TTTCGTCTCCACACCAATCA-----GAGCTGCCACATCTTTGCTGTT 1395  
 Db 94 TrpGlyGlnAspValThrSerValProValThrArgProAlaProGlySerThrThrSer 113  
 QY 1396 ATACAGAGAGTCTACTTTCAGTCCAGAACTTCCTCTGTGAACCCAGCTTGACACA 1455  
 Db 114 ProAlaGlnAspValThrSerAlaProAsp-----ThrArgProAlaLeuGlySer 130  
 QY 1456 GTGACGGGACAGACAGTGTCTA-----CTGACACTTCTTGTCTCCACCTGCT 1506  
 Db 131 ThrAlaProValHisGlyValThrSerAlaProAspThrArg-----ProThr 147  
 QY 1507 ATGCGCTTACCTCCCTCTCAGAGCTCCACCTTCTTATGCGATCAGCATCTCTCT 1566  
 Db 148 LeuGlySerThr-----AlaProProVal----- 155  
 QY 1567 CTGACTGATCAAGGCACACAGATACATGTCGCACTGACACAGACATCTAGTACCAGG 1626  
 Db 156 -----HisGlyValThrSer-----AlaProAsp----- 163  
 QY 1627 CTCACATCCACACAGATGATTCTCGCAATGACCAACTGCGCTCTGGATTTCCAT 1686  
 Db 164 ---ThrArgProThr-----LeuGlySerThrAla 172  
 QY 1687 CCACCT-----GCATCTTCAGATGACAGCGATCAAGTGCAGTGGCGGAGATG 1737  
 Db 173 ProProValHisAsnValThrSerAlaSerGlySerAlaSerGlySerAlaSerThrLeu 192  
 QY 1738 GTCAGACCTAGATGAATGGATCTGTGCACACTCTGCCCCCACTCGAGGTACCAGAG 1797  
 Db 193 ValHisAsnGlyThrSerAlaArgAlaThrThrProAlaSerLysSerThrPro--- 211  
 QY 1798 CTCAGCGAATATGTTCTGCCGATCATTTCTGGAGGATACCATCTCTGTC 1851  
 Db 212 -----PheSerIleProSerHisHisSerAspThrProThrThrLeuThrSer 227  
 QY 1852 -----TCAGCTTACAGTATATACCACT 1875  
 Db 228 HisSerThrLysThrAspAlaSerSerThrHisHisSerThrValSerProLeuThrSer 247  
 QY 1876 AGTTCATGACCAATGCCCCAAGGCGGAGAGCTGTAGTGTCTTCTC-----AGTCTG 1929  
 Db 248 SerAsnHisSerThrSerProGlnLeuSerIleGlyValSerPhePheLeuSerPhe 267  
 QY 1930 CGTGTGTACATGCGCTTCTCCACAGCTTTCACAGCTTTCACAGAGCTCTCGAGTACCGA 1989  
 Db 268 HisIleSerAsnLeuGlnPheAsnSerSerLeuGluAspProSerThrAsnTrpTrpGln 287  
 QY 1990 GCTCTGGACCAACAAATTCACAGCTGCTGTTCCA-----TATCTACGATCCATCTTACA 2046  
 Db 288 GluLeuGlnArgAspIleSerGluLeuIleGluGlnIleTrpLysGlnGlyAspPheLeu 307  
 QY 2047 GGATTAAAGCACTTGAATACTTAATCTTCAGAAACGGGAGTGTGATTGTGAATAGCAA 2106

Db 308 GlyValSerAsnIleLys-----PheArgProGlySerValValValGlnSerThr 324  
 QY 2107 ATCAAGTTTGTAGTCTGTGCGGTATACCTCACCAGGCTGTGCGACGGGCTTTGGAG 2166  
 Db 325 LeuAlaPheArgGluGly-----ThrAsnValHisAspValGluAla 339  
 QY 2167 GATTTT-----CGTTCTGCTGAGCCCAACACTCCATCTCGAAATGACAGCTAC 2217  
 Db 340 GlnPheAsnGlnHisLysThrGluAlaSerArgTyrAsnLeuThrIleSerAspVal 359  
 QY 2218 TCTTCAACATTTGAACCA 2235  
 Db 360 SerValSerAspValPro 365

RESULT 3  
 DMD\_MOUSE  
 ID DMD\_MOUSE STANDARD; PRT; 3678 AA.  
 AC P11531; O35653; Q60703;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Dystrophin.  
 GN DMD.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92253376; PubMed=1579466;  
 RA Bies R.D., Phelps S.F., Cortez M.D., Roberts R., Caskey C.T.,  
 RA Chamberlain J.S.;  
 RT "Human and murine dystrophin mRNA transcripts are differentially  
 expressed during skeletal muscle, heart, and brain development.";  
 RL Nucleic Acids Res. 20:1725-1731(1992).  
 RN [2]  
 RP SEQUENCE OF 1-201 FROM N.A.  
 RX MEDLINE=87273512; PubMed=3607877;  
 RA Koenig M., Hoffman E.P., Bertelson C.J., Monaco A.P., Feener C.,  
 RA Kunkel L.M.;  
 RT "Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and  
 preliminary genomic organization of the DMD gene in normal and  
 affected individuals.";  
 RL Cell 50:509-517(1987).  
 RN [3]  
 RP SEQUENCE OF 120-176 FROM N.A.  
 RX STRAIN=129/J;  
 EX MEDLINE=92182520; PubMed=1543903;  
 RA Maconochie M.K., Brown S.D.M., Greenfield A.J.;  
 RT "sequence analysis of two exons from the murine dystrophin locus.";  
 RL Mamm. Genome 2:64-68(1992).  
 RN [4]  
 RP SEQUENCE OF 300-1390 FROM N.A.  
 RX MEDLINE=88018015; PubMed=3659917;  
 RA Hoffman E.P., Monaco A.P., Feener C.C., Kunkel L.M.;  
 RT "Conservation of the Duchenne muscular dystrophy gene in mice and  
 humans.";  
 RL Science 238:347-350(1987).  
 RN [5]  
 RP SEQUENCE OF 986-1056 FROM N.A.  
 RX STRAIN=C57BL/10; TISSUE=Skeletal muscle;  
 RX MEDLINE=94154933; PubMed=8111539;  
 RA Chamberlain J.S., Phelps S.F., Cox G.A., Maichele A.J.,  
 RA Greenwood A.D.;  
 RT "PCR analysis of muscular dystrophy in mdx mice.";  
 RL Mol. Cell Biol. Hum. Dis. Ser. 3:167-189(1993).  
 RN [6]  
 RP ALTERNATIVE SPLICING.  
 RX STRAIN=C57BL/10; TISSUE=Retina;  
 RX MEDLINE=95360002; PubMed=7633443;  
 RA D'Souza V.N., Nguyen T.M., Morris G.E., Karges W., Pillers D.-A.M.,  
 RA Ray P.N.;

EMBL; M68859; AAB02797.1; -  
EMBL; X58153; CAA41157.1; -  
EMBL; M18025; AAA37530.1; -  
EMBL; U56724; AAB01216.1; -  
EMBL; U15218; AAB87068.1; -  
PIR; S28916; S28916.  
HSPSP; P46939; IQAG.  
MGSD; MG1; 94909; Dmd.  
GO; GO:0045202; C:synaptic junction; IDA.  
InterPro; IPR001589; Actbind\_actinin.  
InterPro; IPR001715; Calponin-like.  
InterPro; IPR002017; Spectrin.  
InterPro; IPR001202; WW\_Rsp5\_WWP.  
InterPro; IPR000433; Znf\_ZZ\_  
Pfam; PF00307; CH; 2.  
Pfam; PF00435; spectrin; 20.  
Pfam; PF00397; WW; 1.  
Pfam; PF00569; ZZ; 1.  
SMART; SM00033; CH; 2.  
SMART; SM00150; SPEC; 21.  
SMART; SM00456; WW; 1.  
SMART; SM00291; Znf\_ZZ; 1.  
PROSITE; PS00019; ACTININ\_1; 1.  
PROSITE; PS00020; ACTININ\_2; 1.  
PROSITE; RS50021; CH; 2.  
PROSITE; PS01159; WW\_DOMAIN\_1; 1.  
PROSITE; PS00020; WW\_DOMAIN\_2; 1.  
PROSITE; PS01357; ZF\_ZZ\_1; 1.  
PROSITE; PS0135; ZF\_ZZ\_2; 1.  
PROSITE; PS0135; ZF\_ZZ\_2; 1.  
Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;

|    |      |   |                                       |                         |     |
|----|------|---|---------------------------------------|-------------------------|-----|
| QY | 173  | TTTCTCCAGTTC  | CAAGGACCAAGATATCTCCATTAA              | CATATACCATTCGTGAACAATAA | 232 |
|    |      |   | :::                                   |                         |     |
| Db | 1965 | PheAlaGlnPheArgLeuAsnPheAlaGlnIleHisThrLeuHisGluGluThrMet     | 1984                                  |                         |     |
| QY | 233  | -----GACATGACACATCCCCCAGAGAATGAAACAACACT                      | 265                                   |                         |     |
|    |      |   |                                       |                         |     |
| Db | 1985 | ValValThrThrGluaspMetProLeuaspValSerTyrlProSerThrrtyrLeuThr   | 2004                                  |                         |     |
| QY | 266  | GAAAGTACTGA   | AAAAATCTCAAAATGTCACATATGAGACGAAATATTC | --- 313                 |     |
|    |      |   | :::                                   |                         |     |
| Db | 2005 | GluileSerHisIleLeuGlnAlaLeuSerGluValAspHisLeuLeuAsnThrProGlu  | 2024                                  |                         |     |
| QY | 314  | -----GATTGGCGAAGCATCGACAAAAGAACATCCGATTTTTC                   | 352                                   |                         |     |
|    |      |   |                                       |                         |     |
| Db | 2025 | LeuCysAlalyeAspPheGluaspLeupPheLysGln                         | ----- 2036                            |                         |     |
| QY | 353  | CCAACGGGGGTTTAAGTCTGTCCACAGGAATCGATGAAACAGATTTTGACAGCTTCAA    | 412                                   |                         |     |
|    |      | :::     :::   |                                       |                         |     |
| Db | 2037 | -----GluGluSerLeuLysAsnIleLysAspAsnLeuGln                     | 2048                                  |                         |     |
| QY | 412  | ----- 412   |                                       |                         |     |
| Db | 2049 | GlnIleSerGlyXrglleaspileIleHlsylsGlysLysThrAlaAlaLeuGlnSerAla | 2068                                  |                         |     |
| QY | 413  | ----GCYTATTATAGATTGAGCTGTGTCCAGACAGCAGTAGTATGGGAA             | --- 454                               |                         |     |
|    |      | :::     :::   |                                       |                         |     |
| Db | 2069 | ThrSerMetGluLysVallysVal                                      | --GlnGluAlaValAlaGlnMetAspPheGlnGly   | 2087                    |     |
| QY | 455  | -----GCATATCGATCTTTCGATCGCATCCCT                              | ----- 484                             |                         |     |

2088 GluLysLeuHisArgMetTyrLysGluArgGlnGlyArgPheAspArgSerValGluLys 2107  
 485 -----GACACAGGGGAATATCAGACTGGGTGCAGCATCTGCCAGCAG 526  
 2108 TrpArgHisPheHisTyrAspMetLysValPheAsnGlnTrpLeuAsnGluValGluGln 2127  
 527 GAGACCTTCCTCCCTTTGCATCTGGAAAAAACCCTTCAGCAATCCCGAGG----- 577  
 2128 -----PhePheLysThrGlnAsnProGluAsnTrpGluHisAlaLysTyr 2143  
 578 -----CACCTGGATCTTCCAGCAGAGATAAAACAGAGAAGTTTCCCTGCACAGAAA 631  
 2144 LysTrpTyrLeuLysGluLeuGlnAspGlyLeuGlnArgGln----- 2158  
 632 GATGAAATATCTGCAGAGAGACATTGGGAGACCTCGTGAACCATGTCTCAATTCAACA 691  
 2159 -----AlaValValArgThrLeuAsnAlaThrGlyGluGluLeuGlnGlnSer 2175  
 692 GC-----AATCTACATTCAAAAGACTTTGGGCAGTATT----- 723  
 2176 SerLysThrAspValAsnLeuGlnGluLysLeuGlySerLeuSerLeuArgTrpHis 2195  
 724 -----CTAAGAAAACCCCTCAGAGAGCAA----- 747  
 2196 AsnIleCysLysGluLeuAlaGluArgLysArgIleGluGlnLysAsnValLeu 2215  
 748 -----ATTCAAGATGTTGCCAAGTC 768  
 2216 SerGluPheGlnArgAspLeuAsnGluPheValLeuTrpLeuGluGlnAlaAspAsnIle 2235  
 769 TCACCTGGGCTTCCCTCTCACTCTGTATGACACCTCTCAATCAATTCAGATTC 828  
 2236 AlaIleThrProLeu-----GlyAspGluGlnLeuLysGlu----- 2248  
 829 ACACCTCAACGACACCAAGATCCCTACCAACAGAAAGAAACAGATTCGCTGTGTGGAG 888  
 2249 GlnLeuGluGlnValLysLeuLeuAlaGluLeuProLeuArgGlnGlyLeuLys 2268  
 889 GAGCAGAGGGTGGAGCTCAGCGTCTCTGTGTAACACAGAGAAGTTCAAGCGCAGCTCGCT 948  
 2269 GlnLeuAsnGluThrGlyGlyAlaValLeuValSerAlaProIleArgProGluGln 2288  
 949 GACTCCAGTCCGCATATTACACGAGAGCTAGCAGGAAGTCCCACTTCAGATGCAAAAG 1008  
 2289 AspLys-----LeuGluLysLysLeuLysGlnThrAsnLeuGlnTrpLys 2304  
 1009 ATATTAAAGAACTTCCA-----GGATCAAAAATCCATGTTTAGATTAGA 1059  
 2305 ValSerArgAlaLeuProGluLysGlnGlyLeuGluValHisLeuLysAspPheArg 2324  
 1060 CCAAAGAAAAGATAGTGGCTCAAGCTCCACAGATGCAACTTACGGCCATCTTTAAG 1119  
 2325 GlnLeuGluGln-----LeuAspHisLeuLeuLeuTrpValSerProIle 2340  
 1120 AGACACAGTGCAGAGCAAAAAGCCCTGCAGTGACCTCTCTGTTGATTCCACAAA 1179  
 2341 ArgAsnGlnLeuGluIleTyrAsnGlnProSerGlnAlaGlyProPheAspLysGlu 2360  
 1180 ATTGAAAGTGAAGAGCTATCATATGACACCATGGAGGACACAGCAACAGAAATCTAT 1239  
 2361 Ile-----GluValThrValHisGly-----LysGln----- 2369  
 1240 CTCACAGCTACAGACCTCAAAAGGCTGATCAGCAAAAGCA-----CTAGAGGAACA 1293  
 2370 -----AlaAspValGluArgLeuLeuSerLysGlyGlnHisLeuTyrLysGluLys 2386  
 1294 TCTTTGGATGTGGGACCAATTCAGTTCATGATGAAATTCGTGGATCACTGCCACCTTT 1353  
 2386 ----- 2386  
 1354 GGTCTCTGACCCCAATCAGAGCTGCCACATCTTTTGTGTTATPACAGAGGATGCTACT 1413

|   |   |  |      |   |  |       |      |
|---|---|--|------|---|--|-------|------|
| D | b |  | 2387 | ---   | ProSerThrGln-                          | ----- | 2390 |
| Q | y |  | 1414 | TTGAGTCCAGAACTTCCTCCTGTGAACCCACCGCTTGAGACAGTG-                    | -----GAC                               | 1461  |      |
| D | b |  | 2391 | :::   | :::                                    |       |      |
|   |   |  |      | ProValIysArgLysLeuGluAspLeuAsgSerGluTrpGlu                        | 2404                                   |       |      |
| Q | y |  | 1462 | GGAGCAGACATGGTCTTA  | -----CCTGCACACTTCCTGGTCTCCA            | 1500  |      |
| D | b |  | 2405 | AlaValasnHisLeuLeuArgGlnLeuArgThrLysGlnProAspArgala               | -----                                  | 2421  |      |
| Q | y |  | 1501 | CCTGCTATGGCCTCACTACCTCCCTGTGCAAGACTCCACCTTCTTTATTGATGCCATCAAGCATC | 1560                                   |       |      |
| D | b |  | 2422 | ProGlyLeuSerThrThrGlyAlaSer                                       | -----AlaSerGlnThr                      | 2434  |      |
| Q | y |  | 1561 | TTCCTCTGCTACTGATCAAGGCACACACAGATACAATGGCCACTGACACGACAATGCTAGTA    | 1620                                   |       |      |
| D | b |  | 2435 | ValThrLeu-  | -----ValThrGlnSerValValThrLysGluThr    | 2449  |      |
| Q | y |  | 1621 | CAAGGCTCACCAATCCCACAGCTGATTTATTCGCAATCAGCAACACTGCCTCTGGGAATT      | 1680                                   |       |      |
| D | b |  | 2450 | SerLysLeuGluMetProSerSer  | -----LeuLeuLeuGluVal                   | 2462  |      |
| Q | y |  | 1681 | TCACATCCACCTGCATCTTCAGATGACAGCCGATCAAGTGGAGATATGCTC               | 1740                                   |       |      |
| D | b |  | 2463 | -----ProalaleuAlaaspPheAsnArgAlatrpThrGluLeuThrAspTrpLeu          | 2479                                   |       |      |
| Q | y |  | 1741 | AGACACCTAGATCAAAATGGATCTCTCTGACACTCTCTGCCCATCTCAGGTACCGAGCTC      | 1800                                   |       |      |
| D | b |  | 2480 | SerLeuLeuAspArgValIleLysSerGlnArgValMetValGlyAspLeuGluAspIle      | 2499                                   |       |      |
| Q | y |  | 1801 | ACGGAATATGTT  | -----TCTGTCCCAAGCATTTCTCTGAGGATACCACCT | 1845  |      |
| D | b |  | 2500 | AsnGluMetIleIleLysGlnLysAlaThrLeuGlnAsp                           | -----LeuGluGlnArgArg                   | 2517  |      |
| Q | y |  | 1846 | CCTGTCTCAGCTTACAGTATATCAACACTAGTCTTATGACATTCATGACATGGCCCCAAC      | 1899                                   |       |      |
| D | b |  | 2518 | Pro- -GlnLeuGlnGluLeuIleThrAlaAlaGlnAsnLeuLysAsnLysThrSer         | 2535                                   |       |      |
| Q | y |  | 1900 | GCCGAGAGCTGGTAGTGTCTTCAGTCTGGCTGTCTCTAAACATGGCCCTCTCCACAGC        | 1959                                   |       |      |
| D | b |  | 2536 | AsnGlnGluAlaArgThrIleThrAspArgIleGluArgIleGlnIleGlnTrpAsp         | 2555                                   |       |      |
| Q | y |  | 1960 | -----CTGTTCAAC  | -----AAGAGCTCT                         | 1977  |      |
| D | b |  | 2556 | GluValGlnGlnGlnLeuGlnAsnArgGlnGlnLeuAsnGluMetLeuLysAspSer         | 2575                                   |       |      |
| Q | y |  | 1978 | CTGAGTACCGAGCTCTGGAGCAACAATTCACACAGCTGCTG                         | -----                                  | 2019  |      |
| D | b |  | 2576 | ThrGlnTrpLeuGluAlaLysGlnGluAlaGlnGlnValIleGlyGlnValArgGlyLys      | 2595                                   |       |      |
| Q | y |  | 2020 | -----GTTCATATCTACGATCCAATCTTACAGGA                                | 2049                                   |       |      |
| D | b |  | 2596 | LeuAspSerTrpLysGluGlyProHishrValAspAlaIleGlnLysLysIleThrGlu       | 2615                                   |       |      |
| Q | y |  | 2050 | TTTTAAGCAACT--GAATACTTAATCTTCAGAAAACGGAGGTGATGTGTG                | -----AAT                               | 2100  |      |
| D | b |  | 2616 | ThrLysGlnLeuAlaLysAspLeuArgGlnArgGlnIleSerValAspValAlaAsnAsp      | 2635                                   |       |      |
| Q | y |  | 2101 | AGCAAAATCAAGTTGCTTGAAGTCTGTGCGGTATTAACCTCACCAAGCGGTGTGACGGGGTC    | 2160                                   |       |      |
| D | b |  | 2636 | LeuAlaLeuLysLeuLeuArgAspTyrSerAlaAspAspThrArgLysValHisMetIle      | 2655                                   |       |      |
| Q | y |  | 2161 | TTGAGAGATTTT  | -----CGTCTGCT                          | 2181  |      |
| D | b |  | 2656 | ThrGluAsnIleAsnThrSerTrpGlyAsnIleHisLysArgValSerGluGlnGluAla      | 2675                                   |       |      |
| Q | y |  | 2182 | GCAGCCCCAACTCCATCTGGAAATAGACAGCTACTCTCTCAACATTCGAACCACTGAT        | 2241                                   |       |      |
| D | b |  | 2676 | AlaLeuGluGluThrHisArgLeuLeuGlnGlnPheProLeuAspLeuGlu               | -----                                  | 2692  |      |
| Q | y |  | 2242 | CAAGCAGATCCCTCGAAGTTCTCGCCTCGCGGGAATTTGCCCAANTGTGTAAGAAC          | ---                                    | 2298  |      |
| D | b |  | 2693 | -----LysPheLeuSerTrpIleThrGluAlaGluThrThrAlaAsnVal                | 2707                                   |       |      |

```

QY 2299 -----GACGAGCTAGGAGCGAGTGTGCG 2325
Db 2708 LeuGlnAspAlaSerArgGlyLeuGluAspSerArgGlyValArgGluLeu 2727
QY 2326 TCGAACCCAGGATATGACAGCGAGGAGCTTGGAC -----GGTCTGGAACCAAGGC 2376
Db 2728 MetLysProTrpGlnAspLeuGlnGlyGluIleGluThrHisThrAspIleTyrHisAsn 2747
QY 2377 CTCTGTGGCCCTGCACAAAGGATCGAGGTCTCCAGGAAAGGAGCTCCATGCAGG 2436
Db 2748 LeuAspGlnAsnGlyGlnLysIleLeuArgSerLeuGluGlySerAspGluAlaProLeu 2767
QY 2437 TTGCCAGATCACTCTGAAATCAAGCATCAAACTAGT -----GTTAAAGATTCCTCAA 2490
Db 2768 LeuGlnArgArgLeuAspAsnMetAsnPhelYsrTrpSerGluLeuGlnLysLysSerLeu 2787
QY 2491 AATCAACAATAATCAAGGTAAATCAAGTAAAGAAATTTCTGAATTTACTGACGTAGATAT 2550
Db 2788 AsnIleArgSerHisLeuGluAlaSerSerAspGlnTrpLysArgLeuHisLeuSerLeu 2807
QY 2551 GAAGAA 2556
Db 2808 GlnGlu 2809

RESULT 4
ID PGCV MOUSE STANDARD; PRT: 3358 AA.
AC Q62059; Q62059; Q9CUU0;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M).
GN CSFG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A. (ISOFORMS V0; V1 AND V2).
RC STRAIN=C57BL/6; and Swiss Webster; TISSUE=Brain;
RX MEDLINE=9512551; PubMed=7822336;
RA Ito K., Shinomura T., Zako M., Ujita M., Kimata K.;
RT "Multiple forms of mouse PG-M, a large chondroitin sulfate
RT proteoglycan generated by alternative splicing.";
RL J. Biol. Chem. 270:958-965 (1995).
RN [2]
SEQUENCE FROM N.A. (ISOFORM V3).
RC STRAIN=C57BL/6;
RX MEDLINE=95181355; PubMed=7876137;
RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
RT without a chondroitin sulfate attachment in region in mouse and human
RT tissues.";
RL J. Biol. Chem. 270:3914-3918 (1995).
RN [3]
SEQUENCE OF 1-1692 FROM N.A. (ISOFORM V1).
RC STRAIN=C57BL/6; TISSUE=Skin;
RX MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tonaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojibori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasaki Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagaehima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,

```

```

RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reid D.J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilmig L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Maki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sakaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
RN [4]
INTERACTION WITH FBLN1.
RP MEDLINE=99329059; PubMed=10400671;
RX Asperger A., Adam S., Kostka G., Timpl R., Heinkegaard D.;
RT "Fibulin-1 is a ligand for the C-type lectin domains of aggrecan and
RT versican.";
RL J. Biol. Chem. 274:20444-20449 (1999).
CC -!- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -!- SUBUNIT: Interacts with FBLN1.
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Comment=Additional isoforms seem to exist;
CC Name=V0;
CC IsoId=Q62059-1; Sequence=Displayed;
CC Name=V1;
CC IsoId=Q62059-2; Sequence=VSP_003087, VSP_003088;
CC Name=V2;
CC IsoId=Q62059-3; Sequence=VSP_003089;
CC Name=V3;
CC IsoId=Q62059-4; Sequence=VSP_003087, VSP_003090;
CC -!- TISSUE SPECIFICITY: V2 is found only in brain.
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 2 link domains.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D16263; BAA03796.1; -.
CC DR EMBL: D28599; -. NOT ANNOTATED_CDS.
CC DR EMBL: D32040; BAA06802.1; -.
CC DR EMBL: AK014525; BAB29411.2; -.
CC DR HSSP: P01132; IEPG.
CC DR MGD: MGI:102889; Cagp2.
CC DR InterPro: IPR000152; Asx_hydroxyl_s.
CC DR InterPro: IPR000742; EGF_2.
CC DR InterPro: IPR001881; EGF_Ca.
CC DR InterPro: IPR006209; EGF-like.
CC DR InterPro: IPR007110; Ig-like.
CC DR InterPro: IPR003599; Ig.
CC DR InterPro: IPR001304; Lectin_C.
CC DR InterPro: IPR000538; Link.
CC DR InterPro: IPR000436; Sushi_SCR_CCP.
CC DR Pfam: PF00008; EGF; 2.

```

[illegible]

Db 1684 rpValSerAspLysSerPheGluGlyValGlyLysGluLysGluAsnGluAspGluGluClyA 1704  
 QY 1208 CCATGAGGAGGACACACAGCAAGCAATCTATCTCAGCTACAGACTCAAAAGGCTGA 1267  
 Db 1704 laValAsnAlaAlaHieGlnGlyGluValArgAlaAlaThrGluArgSerAspHisLeuL 1724  
 QY 1268 TCAGCAAGACACTAGAGAGAAACAATCTTTGGATGTGGGGACAATTCAGTTCACTGATG 1327  
 Db 1724 euLeuThrProGluLeuGluSerSerAsnValAlaSerSerAspLeuAlaThrTrpG 1744  
 QY 1328 AA-----ATTGCTGGATCACTGCCAGCCTTTGGTCTCGACACCAATCAGACCTGCCA 1381  
 Db 1744 luGlyPheIleLeuGluThrThrProThr-----GluSerGluGlyGluMetalaa 1761  
 QY 1382 CATCTTTTGGCTTTATAACAGAGGAT---GCTACTTTGATCCAGAACTTCCTCTGTTG 1438  
 Db 1761 snSerThrProValPheArgGluThrIleGlyValAlaAsnValGluAlaGlnProPheG 1781  
 QY 1439 AACCCGACTTGGACAGTGGACGAGCAGCAGCATGGTCTACCTGACACTTCTTGGTCTC 1498  
 Db 1781 luHisSerSerSerHisProArgValGlnGluGluLeuThrThrLeuSerGlyAsnP 1801  
 QY 1499 CACCTGCTATG-----GCCTCTACCTCCCTGTCAGAGCTC 1534  
 Db 1801 roProSerLeuPheThrAspLeuGlySerGlyAspAlaSerThrGlyMetGlu----- 1818  
 QY 1535 CACCTTTCTTATGCGCATCAGCATCTTCTCTGACTGATCAAGGCACACA----- 1587  
 Db 1819 -----LeuIleThrAlaSerLeuPheThrLeuAspLeuGluSerGluThrLysValL 1836  
 QY 1588 -----GATA 1591  
 Db 1836 ysLysGluLeuProSerThrProSerProSerValGluIleSerSerSerPheGluProT 1856  
 QY 1592 CAATGGCCACTGACCAAGCAATGCTA-----GTACAGGGCTCACCATCCCAACCA 1642  
 Db 1856 hrGlyLeuThrProSerThrValLeuAspIleGluIleAlaGlyValMetSerGlnThrS 1876  
 QY 1643 GTGATATTCTGCAATCAGCACTGCTGCTGGGATTTTCATCCACTGCTGATCTTCAG 1702  
 Db 1876 erGlnLysThrLeuIleSerGlu-----IleSerGlyLysProThrSerGlnS 1892  
 QY 1703 ATGACAGCGCATCAAGTGCAGGTGGGAGATATGGTCAGACACCTA----- 1749  
 Db 1892 erGlyValArgAspLeuThrThrGlyPheProMetGlyGluAspPheSerGlyAspPheS 1912  
 QY 1749 ----- 1749  
 Db 1912 erGluTyThrProThrValSerTyThrThrMetLysGluGluThrValGlyMetGlyGlyS 1932  
 QY 1750 --GATGAAATGATCTGCTGACACTCTGCCCCATCTGAGTACAGAGCTCAGCGAAT 1807  
 Db 1932 exAspAspGluArgValArgAspThrGlnThrSerSerIleProThrThrSerAspA 1952  
 QY 1808 ATGTT---TCGTCCCGATCATTTCTGGAGGATACCACTCTGCTCTCAGCTTTACAGT 1864  
 Db 1952 snIleTyProValProAspSerLysGlyProAspSerThrValAlaSerThrThrAlap 1972  
 QY 1865 ATATCACCAGTCTATGATGACCATGTCCTCCCAAGCTGTTCAACAGAGCTCTCTGG 1921  
 Db 1972 heProTrpGluGluValMetSerSerAlaGluGlySerGlyGluGluLeuAlaSerVala 1992  
 QY 1922 TCAGTCTCGTGTGTGCTTAACATGGCTCTTCCCAAGCTGTTCAACAGAGCTCTCTGG 1981  
 Db 1992 rgSerSerValGlyProValLeuProLeuAlaValAlaSerIlePheSerGlyThrGluSerP 2012  
 QY 1982 ACTACCGAGCTGGAGCAACAATTCACAGCTG----- 2016  
 Db 2012 roTyf-----PheAspGluGluPheGluGluValAlaAlaValThrGluAlaAsnGluA 2030  
 QY 2017 -----CTGGTTCCATATCTAGCATCCATCTTACAGGATTTAAGCACTTGAATAC 2068  
 Db 2030 rgProThrValLeuProThrAlaAlaSerGlyAsnThr-----Vala 2044

QY 2069 TTAACCTCAGAAACGGAGTGTGATTTGTAATAGCAAAATG----- 2109  
 Db 2044 epLeuThrGluAsnGlyTyIleGluValAsnSerThrMetSerLeuAspPheProGlnT 2064  
 QY 2110 -----AAGTTTCTAGTCTGTGCTGCCGTATAACCTCACCAGAGCTGTGCACG 2155  
 Db 2064 hrMetGluProSerLysLeuTrpSerLysProGluValAsnLeuAspLysGlnGluLeG 2084  
 QY 2156 GG-----GTCITGAGAGATTTTCGTT 2176  
 Db 2084 lyArgGluThrValThrLysGluLysAlaGlnGlyGlnLysThrPheGluSerLeuHis 2104  
 QY 2177 CTGCT---CGACCCACACACTCCATCTGGAATAGACAGTACTCTCTCAACATTGAAC 2233  
 Db 2104 erSerPheAlaProGluGlnThrIleLeuGluThrGlnSerLeu-----IleGluT 2121  
 QY 2234 CAGCTGATCAGCAGATCCCTGCAAGTTCTGCGCTGCGCGGAAATTTGCCAATGTGTAA 2293  
 Db 2121 hrGluPheGlnThrSerAspTyThrSerMetLeuThr-----ThrLeuLysThrTyIleT 2139  
 QY 2294 AGAACGACGAGCTGAGAGCGAGTGTGCTGCAACCGAGATATCAGACCGAGGA 2353  
 Db 2139 hrAsnLysGluValGluGlu----- 2145  
 QY 2354 GCTTGGACGCTCTG-----GAACGAGCCTCTGTGGCCCTGGCACAAAGGATGCG 2404  
 Db 2146 -----GluGlyMetSerIleAlaHisMetSerThrProGlyProGlyLysAspLeuG 2164  
 QY 2405 AG----- 2406  
 Db 2164 lySerTyThrThrHisProGluAlaProGlyLysSerHisSerPheSerAlaThrAlaL 2184  
 QY 2407 --GTCTTCCAGGAAAGGAGGATCCATGCTGAGTTCAGATCCTCTGAAATCAGCAT 2464  
 Db 2184 euValThrGluSerGlyAlaAlaArgSerValLeuMetAspSerSerThrGlnGlu----- 2202  
 QY 2465 ACACAACTAGCTGTATAAGTTCCAA-----AATCAACAAATAACA 2506  
 Db 2203 --GluGluSerIleLysLeuPheGlnLysGlyValLysLeuThrAsnLysGluSerAsnA 2222  
 QY 2507 AGGTAATCAGT 2517  
 Db 2222 laAspLeuSer 2225  
 RESULT 5  
 SNEI\_HUMAN  
 ID SNEI\_HUMAN STANDARD; PRT; 8797 AA  
 AC Q9NF91; Q94890; Q9N9P7; Q8TCPI; Q8WWE6; Q8WWW7; Q8WKF6; Q96N17;  
 AC Q9COA7; Q9H525; Q9H526; Q9NS36; Q9NU50; Q9UU06; Q9UU07; Q9ULF8;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Nesprin 1 (Nuclear envelope spectrin repeat protein 1) (Synaptic  
 DE nuclear envelope protein 1) (Syn-1) (Myocyte nuclear envelope protein  
 DE 1) (Wye-1) (Enactin).  
 GN SYNE1 OR MYNE1 OR KIAA0796 OR KIAA1756 OR KIAA1262.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), CHARACTERIZATION, AND  
 RP MUTAGENESIS OF 8758-LEU--CYS-8763.  
 RC TISSUE=Heart, Placenta, Skeletal muscle, Spleen, and Testis;  
 RX MEDLINE=21652858; PubMed=11792814;  
 RA Zhang Q., Skepper J.N., Yang F., Davies J.D., Hegyi L., Roberts R.G.,  
 RA Weissberg P.L., Ellis J.A., Shanahan C.M.;  
 RT "Nesprins: a novel family of spectrin-repeat-containing proteins that  
 RT localize to the nuclear membrane in multiple tissues.";  
 RL J. Cell Sci. 114:4485-4498(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND SUBCELLULAR LOCATION.

CC RX TISSUE=Heart, Spleen, and Testis;  
 CC RX MEDLINE=22296983; PubMed=12408964;  
 CC RT Zhang Q., Ragnauth C., Greener M.J., Shanahan C.M., Roberts R.G.;  
 CC RT "The neprins are giant actin-binding proteins, orthologous to  
 CC RT Drosophila melanogaster muscle protein MSP-300.";  
 CC RL Genomics 90:473-481(2002).  
 CC RN [3]  
 CC RP SEQUENCE FROM N.A. (ISOFORM 4), AND VARIANT GLY-8323.  
 CC RA Braune S., Abraham S., Padmakumar V., Tunggal B., Noegel A.A.,  
 CC RA Korenbaum E.;  
 CC RT "The longest isoform of enaptin/Syne-1, a nuclear envelope associated  
 CC RT protein, binds actin cytoskeleton via the alpha-actinin-like actin-  
 CC RT binding domain.";  
 CC RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
 CC RN [4]  
 CC RP SEQUENCE FROM N.A. (ISOFORMS 8 AND 9).  
 CC RA Zhang Q., Shanahan C.M.;  
 CC RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 CC RN [5]  
 CC RP SEQUENCE FROM N.A.  
 CC RA Almeida J., Clark S., Griffiths C., Lloyd D., Parker A., Smith M.,  
 CC RA Tracey A., Williams S.;  
 CC RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 CC RN [6]  
 CC RP SEQUENCE OF 1-856 FROM N.A.  
 CC CC TISSUE=Kidney;  
 CC RA Gough L., Fan J., Lisa G., Chu S., Winnick S., Beck K.A.;  
 CC RT "Golgi localization of syne-1.";  
 CC RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 CC RN [7]  
 CC RP SEQUENCE OF 28-778 AND 2901-3476 FROM N.A.  
 CC CC TISSUE=Adrenal gland, and Teratocarcinoma;  
 CC RA Niimiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,  
 CC RA Fureya T., Takahashi M., Kikkawa B., Omura Y., Abe K., Kamihara K.,  
 CC RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,  
 CC RA Otaki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
 CC RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
 CC RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,  
 CC RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,  
 CC RA Takemoto M., Ota T., Suzuki Y., Sugano S., Nagahara K., Masuho Y.,  
 CC RA Nagai K., Isoqai T.;  
 CC RT "NED0 human cDNA sequencing project.";  
 CC RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 CC RN [8]  
 CC RP SEQUENCE OF 443-8797 FROM N.A. (ISOFORM 5).  
 CC CC TISSUE=Brain;  
 CC RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;  
 CC RT "Prediction of the coding sequences of unidentified human genes. XIX.  
 CC RT The complete sequences of 100 new cDNA clones from brain which code  
 CC RT for large proteins in vitro.";  
 CC RL DNA Res. 7:347-355(2000).  
 CC RN [9]  
 CC RP SEQUENCE OF 743-8797 FROM N.A. (ISOFORM 6).  
 CC CC TISSUE=Brain;  
 CC RA Ansoorge W., Winkner U., Mewes H.-W., Weil B., Wiemann S.;  
 CC RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 CC RN [10]  
 CC RP SEQUENCE OF 4219-8797 FROM N.A. (ISOFORM 7).  
 CC CC TISSUE=Brain;  
 CC RA Nagase T., Ishikawa K.-I., Kikuno R., Hiroseawa M., Nomura N.,  
 CC RA Ohara O.;  
 CC RT "Prediction of the coding sequences of unidentified human genes. XV.  
 CC RT The complete sequences of 100 new cDNA clones from brain which code  
 CC RT for large proteins in vitro.";  
 CC RL DNA Res. 6:337-345(1999).  
 CC RN [11]  
 CC RP SEQUENCE OF 6922-8797 FROM N.A.  
 CC CC TISSUE=Brain;  
 CC RX MEDLINE=99087487; PubMed=9872452;  
 CC RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,  
 CC RA Tanaka A., Kotani H., Nomura N., Ohara O.;

CC RT "Prediction of the coding sequences of unidentified human genes. XI.  
 CC RT The complete sequences of 100 new cDNA clones from brain which code  
 CC RT for large proteins in vitro.";  
 CC RL DNA Res. 5:277-286(1998).  
 CC RN [12]  
 CC RP REVISIONS.  
 CC CC TISSUE=Brain;  
 CC RX MEDLINE=22158633; PubMed=12168954;  
 CC RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;  
 CC RT "Construction of expression-ready cDNA clones for KIAA genes: manual  
 CC RT curation of 330 KIAA cDNA clones.";  
 CC RL DNA Res. 9:99-106(2002).  
 CC RN [13]  
 CC RP SEQUENCE OF 7631-8797 FROM N.A., AND CHARACTERIZATION.  
 CC RA MEDLINE=21659781; PubMed=11801724;  
 CC RA Mislou J.M.K., Kim M.S., Davis D.B., McNally E.M.;  
 CC RT "Myne-1, a spectrin repeat transmembrane protein of the myocyte inner  
 CC RT nuclear membrane, interacts with lamin A/C.";  
 CC RL J. Cell Sci. 115:61-70(2002).  
 CC RN [14]  
 CC RP SEQUENCE OF 8406-8797 FROM N.A.  
 CC RA Ma F.-R., Zhu L.-F.;  
 CC RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 CC CC -!- FUNCTION: Involved in the maintenance of nuclear organization and  
 CC CC structural integrity. Probable anchoring protein which cheters the  
 CC CC nucleus to the cytoskeleton. Connects nuclei to the cytoskeleton  
 CC CC by interacting with the nuclear envelope and with F-actin in the  
 CC CC cytoplasm.  
 CC CC -!- SUBUNIT: Interacts with MUSK, with F-actin via its N-terminal  
 CC CC domain, and with LMNA in vitro (By similarity).  
 CC CC -!- SUBCELLULAR LOCATION: Type IV membrane protein (Potential). The  
 CC CC largest part of the protein is cytoplasmic, while its C-terminal  
 CC CC outer nuclear membrane. In skeletal and smooth muscles, a  
 CC CC significant amount is found in the sarcomeres.  
 CC CC -!- ALTERNATIVE PRODUCTS:  
 CC CC Event=Alternative splicing; Named isoforms=9;  
 CC CC Name=1;  
 CC CC IsoId=Q8NF91-1; Sequence=Displayed;  
 CC CC Name=2; Synonyms=Beta;  
 CC CC IsoId=Q8NF91-2; Sequence=VSP\_007130;  
 CC CC Name=3; Synonyms=Alpha;  
 CC CC IsoId=Q8NF91-3; Sequence=VSP\_007132, VSP\_007144;  
 CC CC Name=4;  
 CC CC IsoId=Q8NF91-4; Sequence=VSP\_007134, VSP\_007139, VSP\_007140,  
 CC CC VSP\_007144;  
 CC CC Name=5;  
 CC CC IsoId=Q8NF91-5; Sequence=VSP\_007135, VSP\_007136;  
 CC CC Note=No experimental confirmation available;  
 CC CC Name=6;  
 CC CC IsoId=Q8NF91-6; Sequence=VSP\_007137, VSP\_007138;  
 CC CC Note=No experimental confirmation available;  
 CC CC Name=7;  
 CC CC IsoId=Q8NF91-7; Sequence=VSP\_007141, VSP\_007142;  
 CC CC Note=No experimental confirmation available;  
 CC CC Name=8; Synonyms=Beta 2;  
 CC CC IsoId=Q8NF91-8; Sequence=VSP\_007131;  
 CC CC Name=9; Synonyms=Alpha 2;  
 CC CC IsoId=Q8NF91-9; Sequence=VSP\_007133, VSP\_007143, VSP\_007144;  
 CC CC TISSUE SPECIFICITY: Widely expressed. Highly expressed in skeletal  
 CC CC and smooth muscles, heart, spleen, and peripheral blood  
 CC CC leukocytes.  
 CC CC -!- DOMAIN: The Klarsicht domain, which contains a transmembrane  
 CC CC domain, mediates the nuclear envelope targeting.  
 CC CC -!- SIMILARITY: Belongs to the neprin family.  
 CC CC -!- SIMILARITY: Contains 1 actin-binding domain.  
 CC CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.  
 CC CC -!- SIMILARITY: Contains 12 HAT repeats.  
 CC CC -!- SIMILARITY: Contains 1 Klarsicht domain.  
 CC CC -!- SIMILARITY: Contains 31 spectrin repeats.  
 CC CC -!- CAUTION: Ref.5 (CAB55865, CAB55866, CAC16280 and CAC16281)  
 CC CC sequences differ from that shown due to erroneous gene model  
 CC CC prediction.



-!- CAUTION: Ref.7 (BAB71097) sequence differs from that shown due to a chimeric cDNA.  
-!- CAUTION: Ref.14 sequence differs from that shown due to two framehifts in positions 8412 and 8784.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

|       |           |             |                    |
|-------|-----------|-------------|--------------------|
| ENBL; | AY061755; | AAL33798.1; | -;                 |
| ENBL; | AY061756; | AAL33799.1; | -;                 |
| ENBL; | AF495910; | AAN60442.1; | -;                 |
| ENBL; | AF535142; | AAN03486.1; | -;                 |
| ENBL; | AY184203; | AAO27771.1; | -;                 |
| ENBL; | AY184206; | AAO27774.1; | -;                 |
| ENBL; | AL049548; | CAB55865.1; | -;                 |
| ENBL; | AL049548; | CAB55866.1; | -;                 |
| ENBL; | AL078582; | CAB87586.1; | -;                 |
| ENBL; | AL136079; | -;          | NOT ANNOTATED CDS. |
| ENBL; | AL138832; | CAC16280.1; | ALT_SEQ.           |
| ENBL; | AL138832; | CAC16281.1; | ALT_SEQ.           |
| ENBL; | AL357081; | -;          | NOT ANNOTATED CDS. |
| ENBL; | AL450401; | -;          | NOT ANNOTATED CDS. |
| ENBL; | AL589963; | -;          | NOT ANNOTATED CDS. |

ment Scores:

|                   |        |               |      |
|-------------------|--------|---------------|------|
| No.:              | 0.117  | Length:       | 8797 |
| :                 | 122.50 | Matches:      | 173  |
| nt Similarity:    | 38.04% | Conservative: | 145  |
| Local Similarity: | 20.69% | Mismatches:   | 316  |
| Match:            | 2.60%  | Indels:       | 202  |
|                   | 1      | Gaps:         | 41   |

US-10-007-270-1 (1-3330) X SNE1 HUMAN (1-8797)

[illegible]

|    |      |   |      |
|----|------|---|------|
| QY | 1838 | ATACCACTCTCTCAGCTTTACAGTATATACCACTAGTCTTATGACCATTTGCCCA       | 1897 |
| Db | 6812 | ---GlnSerAlaIysAspArgLeuGluPheTrpThrGlnGlnSerValThrVal---     | 6830 |
| QY | 1898 | AGGCCCGAGACTGGTA-----GTGTTCTTC-----                           | 1923 |
| Db | 6830 | lndGluLeuGluMetValArgAspHisLeuAsnAlaPheLeuGluPheSerLysGluValA | 6850 |
| QY | 1924 | -----AGTTCGCTGTTGTGAATCAGCGCTTCCAAAGACCTCTTCACACAGA           | 1972 |
| Db | 6850 | spAlaGlnSerSerLeuLysSerValLeuSerThrGlyAsnGlnLeuLeuArgLeuL     | 6870 |
| QY | 1973 | GCTCTCTG-----GAGTACCGAGCTCTGGAGCAACAATTCACAC                  | 2011 |
| Db | 6870 | ylsLysValAspThrAlaThrLeuArgSerGluLeuSerArgIleAspSerGlnTrpThrA | 6890 |
| QY | 2012 | AGCTCTG-----GTTCCATATCTACATCCAATCTTACAGGATTTTACCAACTTGAA      | 2065 |
| Db | 6890 | spLeuLeuThrAsnIleProAlaValGlnGluLysLeuHisGlnLeu---GlnMetAspL  | 6909 |
| QY | 2066 | TACTTAACCTCAGAAACGGG-----AGTG                                 | 2089 |
| Db | 6909 | ysLeuProSerArgHisAlaIleSerGluValMetSerTrpIleSerLeuMetGluSerV  | 6929 |
| QY | 2090 | TGATTTGTGAATAGCAAAATGAAGTTTCTAGCTCTCCGCTATACCTCACCAGGCTG      | 2149 |
| Db | 6929 | alIleGlnLysAspGluAspAsnIleLysAsnSerIleGlyTr-----LysAlaI       | 6946 |
| QY | 2150 | TGCACGGGCTTTGGAGGATTTTCGTTCTCTGCACGCCCAACACTCCATCTCGAAATAG    | 2209 |
| Db | 6946 | lHisGluTrpLeuGlnLysTrpLysGly-----PhelysIleAspIleA             | 6961 |
| QY | 2210 | ACAGCTACTCTCTCAACATGTGAACCACTGATCAACAGCAGATCCTCGAAGTTCTCGCCT  | 2269 |
| Db | 6961 | snCysLysGlnLeuThrValAspPheValAsnGlnSer-----ValL               | 6975 |
| QY | 2270 | GCGCGGAATTTGCCCATGTGTAAAGACGACGCACTGAGGAAGCGGAGTGTCTGCTGCA    | 2329 |
| Db | 6975 | eGlnIleSerSerGlnAspValGlnSerLysArgSerAspLysThrAspPheA----     | 6993 |
| QY | 2330 | AACCAAGATATGACAGCCAGCGGAGCTCGACGCTGTG3AACCAAGCCTCTGTGGCCCTG   | 2389 |
| Db | 6994 | -----GluGlnLeuGlyAlaMetAsnLysSerTrpGlnIleLeuGlnGlyLeuV        | 7010 |
| QY | 2390 | GCACAAAGGAATGCAGGCTCTCCAGGGAAGGAGCTCCATCGAGTTGCCAGATCAT       | 2449 |
| Db | 7010 | alThrGluLysIleGlnLeuLeuGluGly-----LeuLeuGlnSerTrpS            | 7025 |
| QY | 2450 | CTGAAATCAAGCATACAAACTAGTGTTAA-----AAGTTCACAAATCAAC            | 2497 |
| Db | 7025 | erGlu-----TyrGluAsnAsnValGlnCysLeuLysThrTrpPheGluThrGlnG      | 7042 |
| QY | 2498 | AAATACCAAGGTAATCAGTAAAGAAATCTGAAATTTACTGACCTCGATATGAGAAT      | 2557 |
| Db | 7042 | luLysArgLeuLysGlnGlnHisArgIleGlyAspGlnAlaSerValGlnAsnAlaLeuL  | 7062 |
| QY | 2558 | TTAACCATCAAGATGGGAGGAATTTAAACCTGAAA                           | 2595 |
| Db | 7062 | ysAspCysGlnAspLeuGluAspLeuIleLysAlaLys                        | 7074 |

## RESULT 6

|            |   |
|------------|---|
| ACIN_HUMAN |   |
| ID         | ACIN_HUMAN  |
| AD         | STANDARD; PRT; 1341 AA.   |
| AC         | Q9UKV3; O75158; Q9UG91; Q9UKV1; Q9UKV2;                           |
| DT         | 28-FEB-2003 (Rel. 41, Created)                                    |
| DT         | 28-FEB-2003 (Rel. 41, Last sequence update)                       |
| DE         | Apoptotic chromatin condensation inducer in the nucleus (Acinus). |
| GN         | ACINUS OR KIAA0670.   |
| OS         | Homo sapiens (Human).   |
| OC         | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| OC         | Mammalia; Eutheria; Plrimates; Catarrhini; Homnidae; Homo.        |

NCBI\_TaxID=3606;  
[1] SEQUENCE FROM N.A. (ISOFORM 1; 2 AND 3), PARTIAL SEQUENCE, FUNCTION,  
RP MED AND MUTAGENESIS OF ASP-1093  
RX MEDLINE=99418558; Pubmed=10490025;  
RA Sahara S., Aoto M., Eguchi Y., Imamoto N., Yoneda Y., Tsujimoto Y.;  
RT "Acinus is a caspase-3-activated protein required for apoptotic  
RL chromatin condensation.";  
RN Nature 401:168-173(1999).  
[2] SEQUENCE FROM N.A. (ISOFORM 4).  
RP TISSU=Uterus;  
RC Wambutt R., Heubner D., Mewes H.-W., Gassenhuber J., Wiemann S.;  
RA Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
[3] SEQUENCE FROM N.A. (ISOFORM 3).  
RP TISSU=Petal brain;  
RC Li W.B., Gruber C., Jessee J., Polayes D.;  
RA "Full-length cDNA libraries and normalization.";  
RT Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
[4] SEQUENCE OF 56-1341 FROM N.A. (ISOFORM 1).  
RP TISSU=Brain;  
RC MEDLINE=98403880; Pubmed=9734811;  
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,  
RA Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. X.  
RP The complete sequences of 100 new cDNA clones from brain which can  
RL code for large proteins in vitro.";  
RN DNA Res. 5:169-176(1998).  
CC -1- FUNCTION: Induces apoptotic chromatin condensation after  
CC activation by CASP3.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=4;  
CC Name=1; Synonyms=L;  
CC IsoId=Q9UKV3-1; Sequence=Displayed;  
CC Name=2; Synonyms=S;  
CC IsoId=Q9UKV3-2; Sequence=VSP\_004028;  
CC Name=3; Synonyms=S;  
CC IsoId=Q9UKV3-3; Sequence=VSP\_004026, VSP\_004029;  
CC Name=4;  
CC IsoId=Q9UKV3-4; Sequence=VSP\_004027;  
CC -1- TISSUE SPECIFICITY: Ubiquitous.  
CC -1- PTM: Undergoes proteolytic cleavage; the processed form is active,  
CC contrary to the uncleaved form.  
CC -1- SIMILARITY: Contains 1 SAP domain.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstations  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcement/>  
CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

|    |   |
|----|---|
| CC | EMBL; AF124726; AAD56724.1; -   |
| DR | EMBL; AF124727; AAD56725.1; -   |
| DR | EMBL; AF124728; AAD56726.1; -   |
| DR | EMBL; AL050382; CAB43681.1; -   |
| DR | EMBL; BX247975; CAD62309.1; -   |
| DR | EMBL; AB014570; BAA31645.2; -   |
| DR | GenBank; HGNC:17066; ACINUS.  |
| DR | MIM; 604562; -  |
| DR | GO; GO:0005634; C:nucleus; IDA.   |
| DR | GO; GO:0016887; F:ATPase activity; NAS.                                     |
| DR | GO; GO:0019899; F:enzyme binding; NAS.                                      |
| DR | GO; GO:0003676; F:nucleic acid binding; NAS.                                |
| DR | GO; GO:0030263; F:apoptotic chromosome condensation; IDA.                   |
| DR | GO; GO:0030219; F:erythrocyte differentiation; IEPI.                        |
| DR | GO; GO:0045657; P:positive regulation of monocyte differentiation. .; IEPI. |
| DR | InterPro; IPR003034; SAP.   |
| DR | Pfam; PF02037; SAP; 1.  |

|  |  |
|--|--|
| DR   | SMART; SM00513; SAP; 1.  |
| DR   | PROSITE; P850800; SAP; 1.  |
| KW   | Apoptosis; Nuclear protein; Alternative splicing.                          |
| FT   | DOMAIN 72 106  |
| FT   | SAP.   |
| FT   | DOMAIN 142 442   |
| FT   | GLU-RICH.  |
| FT   | DOMAIN 573 676   |
| FT   | SER-RICH.  |
| FT   | DOMAIN 1114 1131   |
| FT   | PRO-RICH.  |
| FT   | DOMAIN 1132 1341   |
| FT   | ARG/ASP/GLU/LYS-RICH.  |
| FT   | CLEAVAGE (BY CASPASE-3).   |
| FT   | Misling (in isoform 2).  |
| FT   | /FtId=VSP 004025.  |
| FT   | Missing (in isoform 3).  |
| FT   | /FtId=VSP 004026.  |
| FT   | Missing (in isoform 4).  |
| FT   | /FtId=VSP 004027.  |
| FT   | GSPKKCEAEPPAQTQPTSTQTSHPSESRHHTV   |
| FT   | --> MSPADRCRANTIEPATITSLALLQLDQSRTGRL                                      |
| FT   | P (in isoform 2).  |
| FT   | /FtId=VSP 004028.  |
| FT   | SERIHHTV --> MLESKEG (in isoform 3).                                       |
| FT   | /FtId=VSP 004029.  |
| FT   | D-2A: ABOLISHES CLEAVAGE BY CASP3 AND                                      |
| FT   | CHROMATIN CONDENSATION ACTIVITY.   |
| FT   | Q -> H (IN REF. 4).  |
| FT   | CONFLICT 139 139   |
| FT   | SEQUENCE 1341 AA; 151887 MW; 8FE286681F83AB5C CRC64;                       |
| SO   |  |
| Alignment Scores:                              |  |
| Pred. No.:                                     | 0.083  |
| Score:   | 151.00   |
| Percent Similarity:                            | 33.21%   |
| Best Local Similarity:                         | 19.75%   |
| Query Match:                                   | 2.57%  |
| DB:  | 1  |
| US-10-007-270-1 (1-3330) x ACIN_HUMAN (1-1341) |  |
| QY   | 314 GATTTCGAAGCATCGAAACAAAAGATCCGCATTTTCCCACG-----GGGCTT 364               |
| Db   | :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: |
| QY   | 110 ASNLEUGLNLySHISserThrProHisAlaPheGlnProAsnSerGlnIleGlyGlu 129          |
| Db   | :::::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::    |
| QY   | 365 AAAGTCTGTGCCACAGGAATCCATGAAACAGATTTAGACAGCTCTTCAGCTTATTATAGA 424       |
| Db   | :::::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::    |
| QY   | 130 GLUwetsSerGlnAnsnSerPhellellysGlnTyLeuGlulysGlnGluLeuLeuArg 149        |
| Db   | :::::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::    |
| QY   | 425 TTGAGAGTGTCAGGAAGCAGTATGGCAAGCATATCGGATCTTCTTCGATCGCATCCCT 484         |
| Db   | :::::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::    |
| QY   | 150 GlnArGLeuGluArgGluAla----- 157   |
| QY   | 485 GACACAGGGGAATATCAGCATGGCTGAGCATTCGCCACGAGGAGACTTCTGCTCTTTT 544         |
| Db   | :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::      |
| QY   | 158 GluAlaAlaGluLeuGluGlu-----AlaserAlaGluSerGluaspGluMetIleHis 175        |
| Db   | :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::      |
| QY   | 545 GACATTTGA-----AAAACTTCAGCAATTTCCAGACGACCATCTGGAT 586                   |
| Db   | :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::      |
| QY   | 176 ProGluGlyValAlaSerLeuLeuProProappPheGlnSerSerLeuGluArgProGlu 195       |
| Db   | :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::      |
| QY   | 587 CTCTCCAGCMGAGANATAAACACAGAGAAGTTTCCCTGCAGAAAAGATCAAATATCTGCA 646       |
| Db   | :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::      |
| QY   | 196 Leu-----GluLeuSerArgHiserProArgLysserSerSilerGlu 211                   |
| QY   | 647 GAGAGACATTTGGGAGAGCTGGTGAACCATTTGCATTTCAACAGCAATCTCACATTCA 706         |
| Db   | :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::      |
| QY   | 212 GluLySGlyAspSerAspAsp-GluLyssProArgLyssGlyGluArgGlySerArgVa 231        |
| Db   | :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::      |
| QY   | 707 AAGA-----CTTGGCAGTATTCTTAAGNAACCCCTCAAGAGCAAA 748                      |
| Db   | :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::      |
| QY   | 231 lArgGlnAlaArgAlaAlaLyseulerSerGluGlySerGlnProAla-GluGluGluGlu 251      |
| Db   | :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::      |
| QY   | 749 TTCAGATGTTGCCAAGCTCTCACTTGGGCCCTTCCCTCTCACTCTCTGATGACACCCTCC 808       |
| Db   | :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::      |
| QY   | 251 spGInGlu-----ThrProSerArgAsnLeu- 259                                   |
| QY   | 809 TCAATGAAATTCGATATACACTCAACGACACAAGATGCTCTACAGAAAGAGAAA 868             |

Db 536 euAlaLeuAlaLysGlyLeuThrGluGluCysLeuLysGlnProSerLeuGluGlnLysG 556  
 QY 1801 -----AGCAATATGTTCTGTCAGATCATTTCTGAGGATACCACTCTG 1849  
 Db 556 luGlyArgAlaSerHisThrLeuLeuProSerHisArgLeuLysGlnSerAlaAspS 576  
 QY 1850 TCTCAGCTTTACAGTATATACCACTAGTTCTATGACCTGCCCCCAAGGCCGAGAC 1909  
 Db 576 erSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerArg 594  
 QY 1910 TGCTAGTGTCTCAGTCTGCTGTTGCTTAATGCTGCTTCTCCACCACTGTTCAACA 1969  
 Db 595 -----SerProAsp-----SerS 599  
 QY 1970 AGAGCTCTCTGAGTACGAGCTCTGAGCAACAATTCACAGCTGCTGTTCCATATC 2029  
 Db 599 erGlySerArgSerHisSerProLeuArgSerLysGlnArgAspValAlaGlnAlaArgT 619  
 QY 2030 TAGCATCCACTTACAGGATTTAGCAACTTGAATATCTTAATCTTCAAGACGGAGTG 2089  
 Db 619 hrHisAlaAsnProArgGlyArgProLys-----MetGlySerArgSerThrSerG 636  
 QY 2090 TGATTGTGAATAGCAAAATGAAAGTTTCTAAGTCTGTCGCGTATACCTCAACAGGCTG 2149  
 Db 636 luSerArgSerArgSerArgSerArgSerArgSerAlaSerSerAsnSerArgLysSerL 656  
 QY 2150 TGACAC---GGGGTCTGAGGATTTCTGCTGCTGCGACCACTCCATCTCGAAA 2206  
 Db 656 euSerProGlyValSerArgSerSerTh 666  
 QY 2207 TAGCAGCTACTCTCTCAACATTTGAACAGCT-----GATCAGCAGATCTCTGCA 2257  
 Db 667 -----SerTyThrGluThrLysAspProSerSerGlyGlnGluValAlaThrPro 683  
 QY 2258 AGTTCTGCTGGCGGAATTTGCCAATCTGTTAAAGAACGAAACGAGCTAG 2310  
 Db 684 -----ProValProGlnLeuGlnValCysGluProLysGluArgThrSerThrSerS 701  
 QY 2311 -----GAAGCGAGTCTGCTGCAACACCGATATGACGCCAGCGAGGAGCTCTG 2359  
 Db 701 erSerSerValGlnAlaArgLeuSerGlnProGluSerAlaGluLysHisValThrG 721  
 QY 2360 AGGFTCTGGAACAGGCTCTGTCGCTGGCACAAGGATGCGAGTCTTCCACGGAA 2419  
 Db 721 InArgLeuGlnProGluArgGlySerPro-----LysLysCysGluAlaGluGluAlaG 739  
 QY 2420 AGGAGTCTCATGCGAGTTGCCAGATCATCTCTGAAAATCAA 2460  
 Db 739 luProProAlaAlaThrGlnProGlnThrSerGluThrGln 752  
 RESULT 7  
 ID RTN4 RAT  
 AC Q3UK1; Q3UK10; Q9R009; Q9WUE9; Q9WUFO; PRT; 1163 AA.  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 10-OCT-2003 (Rel. 41, Last sequence update)  
 DE Reticulon 4 (Neurite outgrowth inhibitor) (Nogo protein) (Foccen)  
 DE (Glut4 vesicle 20 kDa protein).  
 GN RTN4 OR NOGO.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OC NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 3), AND PARTIAL SEQUENCE.  
 RC STRAIN-Sprague-Dawley; TISSUE=Adipocyte;  
 RX MEDLINE=99249816; PubMed=10231557;  
 RA Morris N.J., Rees S.A., Neveu J.M., Lane W.S., Lienhard G.E.;  
 RT "Cloning and characterization of a 22 kDa protein from rat adipocytes:  
 RL a new member of the reticulon family.";  
 RL Biochim. Biophys. Acta 1450:68-76(1999).  
 EN [2]

RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).  
 EX MEDLINE=20129258; PubMed=10667796;  
 RA Chen M.S., Huber A.B., Van der Haar M.E., Frank M., Schnell L.,  
 RA Spillmann A.A., Christ F., Schwab M.E.;  
 RT "Nogo-A is a myelin-associated neurite outgrowth inhibitor and an  
 antigen for monoclonal antibody IN-1.";  
 RL Nature 403:434-439(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).  
 RC STRAIN-Wistar Kyoto; TISSUE=Vascular smooth muscle;  
 RA Ito T., Schwartz S.M.;  
 RT "Cloning of a member of the reticulon gene family in rat: one of two  
 minor splice variants.";  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP FUNCTION.  
 EX MEDLINE=22033691; PubMed=12037567;  
 RA GrandPre T., Li S., Strittmatter S.M.;  
 RT "Nogo-66 receptor antagonist peptide promotes axonal regeneration.";  
 RL Nature 417:547-551(2002).  
 CC -!- FUNCTION: Potent neurite outgrowth inhibitor which may also help  
 block the regeneration of the nervous central system in adults (By  
 similarity).  
 CC -!- SUBUNIT: Binds to RTN4R. Interacts with Bcl-xL and Bcl-2 (By  
 similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Anchored to the  
 membrane of the endoplasmic reticulum through 2 putative  
 transmembrane domains (By similarity).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=4;  
 CC Name=1; Synonyms=Nogo-A, NI-220-250;  
 CC IsoId=Q3UK1-1; Sequence=Displayed;  
 CC Name=2; Synonyms=Nogo-B, Foccen-M1;  
 CC IsoId=Q3UK1-2; Sequence=VSP\_005658;  
 CC Name=3; Synonyms=Nogo-C, VP20;  
 CC IsoId=Q3UK1-3; Sequence=VSP\_005656, VSP\_005657;  
 CC Name=4; Synonyms=Foccen-M2;  
 CC IsoId=Q3UK1-4; Sequence=VSP\_005659;  
 CC -!- TISSUE SPECIFICITY: Isoforms 1, 2 and 3 are present in optic  
 nerve, spinal cord and cerebral cortex. Isoforms 1 and 2 are  
 present in dorsal root ganglion, sciatic nerve and PC12 cells  
 after longer exposure. Isoforms 2 and 3 are detected in kidney,  
 cartilage, skin, lung and spleen. Isoform 3 is expressed at high  
 level in skeletal muscle. In adult animals isoform 1 is expressed  
 mainly in the nervous system.  
 CC -!- SIMILARITY: Contains 1 reticulon domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF051335; AAF01564.1; -  
 DR EMBL; AJ242961; CAB71027.1; -  
 DR EMBL; AJ242962; CAB71028.1; -  
 DR EMBL; AJ242963; CAB71029.1; -  
 DR EMBL; AF132045; AAD31019.1; -  
 DR EMBL; AF132046; AAD31020.1; -  
 DR GO; GO:0030176; C:integral to endoplasmic reticulum membrane; IDA.  
 DR GO; GO:0005635; C:nuclear membrane; ISS.  
 DR GO; GO:0005515; F:protein binding; ISS.  
 DR GO; GO:0019987; F:negative regulation of anti-apoptosis; ISS.  
 DR GO; GO:0030517; F:negative regulation of axon extension; ISS.  
 DR InterPro; IPR003388; Reticulon.  
 DR Pfam; PF02453; Reticulon; 1.  
 DR PROSITE; PS50845; RETICULON; 1.  
 DR Endoplasmic reticulum; Alternative splicing; Transmembrane.  
 KW DOMAIN 1 989  
 KW CYTOPLASMIC (Potential).  
 FT TRANSMEM 990 1010  
 FT DOMAIN 1011 1104  
 FT LUMENAL (Potential).

|  |   |          |               |                        |  |
|--|---|----------|---------------|------------------------|--|
| FT   | TRANSMEM  | 1105     | 1125          |                        | POTENTIAL.                               |
| FT   | DOMAIN  | 1126     | 1163          |                        | CYTOPLASMIC (Potential).                 |
| FT   | DOMAIN  | 976      | 1163          |                        | RETICULON.                               |
| FT   | DOMAIN  | 33       | 46            |                        | POLY-GLU.                                |
| FT   | DOMAIN  | 73       | 76            |                        | POLY-ALA.                                |
| FT   | DOMAIN  | 140      | 145           |                        | POLY-PRO.                                |
| FT   | VARSPLIC  | 1        | 964           |                        | Missing (in isoform 3).                  |
| FT   | VARSPLIC  | 965      | 975           |                        | /FTID=VSP 005656.                        |
| FT   |   |          |               |                        | AVLSAELSKTS -> MDQKKHKDK (in isoform 3). |
| FT   |   |          |               |                        | /FTid=VSP 005657.                        |
| FT   | VARSPLIC  | 173      | 975           |                        | Missing (in isoform 2).                  |
| FT   |   |          |               |                        | /FTid=VSP 005658.                        |
| FT   | VARSPLIC  | 192      | 975           |                        | Missing (in isoform 4).                  |
| FT   |   |          |               |                        | /FTid=VSP 005659.                        |
| FT   | CONFLICT  | 1130     | 1131          |                        | MISSING (IN REF. 3; AAD31020).           |
| SQ   | SEQUENCE  | 1163 AA; | 126386 MW;    | 8CB894E0954F0B6 CRC64; |  |
| <br>Alignment Scores:                            |   |          |               |                        |  |
| Pred. No.:                                       |   | 0.0918   | Length:       | 1163                   |  |
| Score:   |   | 150.00   | Matches:      | 167                    |  |
| Percent Similarity:                              |   | 34.62%   | Conservative: | 121                    |  |
| Best Local Similarity:                           |   | 20.07%   | Mismatches:   | 287                    |  |
| Query Match:                                     |   | 2.55%    | Indels:       | 257                    |  |
| DB:  |   | 1        | Gaps:         | 42                     |  |
| <br>US-10-007-270-1 (1-3330) x RTN4_RAT (1-1163) |   |          |               |                        |  |
| QY   | 510 TCAGCATCTGCCACGACGACCTTCGCCTCTTGACATTGGAAAAAACTTCAGCAATT        | 5699     |               |                        |  |
| Db   | 204 ThrValSerGlyGlnGluAspPheProSerValLeuLeuGluThrAlaLaSerLeu        | 223      |               |                        |  |
| QY   | 570 CCCAGG-----AGCAGCTGCATCTTC-----                                 | 590      |               |                        |  |
| Db   | 224 ProSerLeuSerProLeuSerThrValSerPheLysGluHisGlyTyrLeuGlyAsnLeu    | 243      |               |                        |  |
| QY   | 591 -----TCACGACGAGATAAACACGAGAAGTTCCCTGCACAGAAAAGATGAATAT          | 641      |               |                        |  |
| Db   | 244 SerIalValSerSerGluGlyThrIleGluThrLeuAsnGluAlaSerLysGlu          | 263      |               |                        |  |
| QY   | 642 CTGACGACGACATCGGAGAGCTGTGTGAACCAT-----TGTCAATTCACAGCAA          | 695      |               |                        |  |
| Db   | 264 LeuProGluArgAlaThrAsnProPheVal--AsnArgAspLeuAlaGluPhe---SerGl   | 282      |               |                        |  |
| QY   | 696 TCTACATTTCAAGACTTGGGCAGT-----                                   | 720      |               |                        |  |
| Db   | 282 uLeuGluTyrSerGluMetGlySerPheLysGlySerProLysGlyGluSerAlaI        | 302      |               |                        |  |
| QY   | 721 -----ATTCTAAGAAAACTTCAGAGACGCAAT                                | 749      |               |                        |  |
| Db   | 302 eLeuValGluAsnThrLysGluGluValIleValArgSerLysAspLysGluAspLeuVa    | 322      |               |                        |  |
| QY   | 750 TCAAGATGTTCCACAGCTCTCACTTGGCGCTTCCTCTCACTCTCTGATGACACCCTCTCT    | 809      |               |                        |  |
| Db   | 322 iCysSerAlaAlaLeuHisSerProGlnGluSerProValGlyLysGluAspArgValVa    | 342      |               |                        |  |
| QY   | 810 CAAT--GAATTCCTCGATAATACACTCAACGACACCAAGATG-----CCTAC            | 854      |               |                        |  |
| Db   | 342 lSerProGluLysThrMetAspIlePheAsnGluMetGlnMetSerValAlaIaProVa     | 362      |               |                        |  |
| QY   | 855 AACAGAAAAGACAAGATTCGCTGTGTTCAG-----                             | 888      |               |                        |  |
| Db   | 362 largGluGluTyraAlaAspPheLysProPheGluGlnAlaIaTrpGluValLysAspThrTy | 382      |               |                        |  |
| QY   | 889 -GAGCAGAGGGTGGAGCTCAGCGTCTCTCTGTAACACAGAGTTCAGGCGACAG-----      | 942      |               |                        |  |
| Db   | 382 rGluGlySerArgAspValLeuAlaAlaargAlaAsnValGluSerLysValAspArgLy    | 402      |               |                        |  |
| QY   | 943 ----CTCGCTGACTCCAGTCCCATTATACCAGGAGTAGCAGGMAAGTCCCCAATCTCA      | 998      |               |                        |  |
| Db   | 402 sCysLeuGluAspSer-----   | 409      |               |                        |  |
| QY   | 999 GATGCAAAAGATATTTAAGAACTTCACGATTCAAAAAATCAATGTGTAGGATTTAG        | 1058     |               |                        |  |

```

Db 723 pLeuPheSerAspSerIleProGluValProGlnThrGlnGluAlaVal----- 741
QY 1923 CAGTCTCGGTGTCTAACTGCTTCTCCACAGCCTGTTCCACAGAGCTCTTCGGA 1982
Db 742 -----MetLeuMetLysGluSerLeuThrGluValSerGluThrVa 755
QY 1983 GTACCGAGCTCTGGAGCAACAATTCACA-----CAGCTGCTGGTTCATATCT 2030
Db 755 lAlaGlnHisGluGluArgLeuSerAlaSerProGlnGluLeuGlyLysProTyrLe 775
QY 2031 ACGATCC-----AATCTACAGATTAAAG----- 2055
Db 775 uGluSerPheGlnProAsnLeuHisSerThrLysAspAlaAlaSerAsnAspIleProTh 795
QY 2056 -----CAACTGCAATCTTAC-----TTGAG 2078
Db 795 xLeuThrLysLysGluLysLysSerLeuGlnMetGluGluPheAsnThrAlaIleTy-Se 815
QY 2079 AAACGGAGGTGATGTGAATAGCAAAATGAAGTTTGCTAAGTCTGTGCGTATACCT 2138
Db 815 rAsnAspAspLeuLeuSerSerLysGluAspLysLysGluSerGluThrPheSerAs 835
QY 2139 CACCAAGCTGTGCACGGGTCTTGAGGATTTCGTT----- 2178
Db 835 pSerSerProIleGlu-----IleIleAspGluPheProThrPheValSerAlaLysAspAs 854
QY 2179 -----GCTGCAGCCCAACACTCCATCTGGAATAGACAGCTACTCT-----CT 2222
Db 854 pSerProLysLeuAlaLysGluTyThrAspLeuGluValSerAspLysSerGluIleAl 874
QY 2223 CAACATTGAACAGCTGATCAAGCAGATCCCTGCAAGTTCTCTGCGCTGCGGC---GAATT 2279
Db 874 aAsnIleGlnSerGlyAlaAspSerLeuProCysLeuGluLeuProCysAspLeuSerPh 894
QY 2280 TGCCCAATGT-----GTAAAGAACGAACGAGTCAAGAGCGGA 2318
Db 894 eLysAsnIleTyProLysAspGluValHisValSerAspGluPheSerGluAsnAspSe 914
QY 2319 GTGTGCTGCAACACAGGATATGACAGCGAGGCGCTGCGAGCTCGACACAGCGCT 2378
Db 914 rSerValSerLysAlaSerIle---SerProSerAsnValSerAlaLeuGlu----- 930
QY 2379 CTGTGGCTGCGCAACAGGAATCGAGGTCTCCAGGAGGAAGGA-----GC 2426
Db 931 -----ProGlnThrGluMetGlySerIleValLysSerLysSerLeuThrLysGluAl 948
QY 2427 TCCATGCAAGTGTCCAGATCACTCTGGAATCAA 2460
Db 948 aGluLysLysLeuProSerAspThrGluLysGlu 959
RESULT 8
BPAL_MOUSE
ID BPAL_MOUSE STANDARD; PRT; 7389 AA.
AC Q91ZU6; Q91ZU7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Bullous pemphigoid antigen 1, isoforms 1/2/3/4 (BPA) (Hemidesmosomal plaque protein) (Dystonia musculorum protein) (Dystonin).
GN BPAG1 OR DST.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND TISSUE SPECIFICITY.
RX STRAIN=BALE/c; TISSUE=Muscle, and Neuron;
RX MEDLINE=21405767; PubMed=11514586;
RA Leung C.L., Zheng M., Prater S.M., Liem R.K.H.;
RT "The BPAG1 locus: alternative splicing produces multiple isoforms with distinct cytoskeletal linker domains, including predominant isoforms in neurons and muscles."
RL J. Cell Biol. 154:691-697 (2001).

```

```

RN [2]
RP SEQUENCE OF 6693-7389 FROM N.A. (ISOFORMS 3 AND 4).
RC STRAIN=C57BL/6J; TISSUE=Fetal skin, and Fetal spinal cord;
RX MEDLINE=22354689; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikiado I., Oeato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Mateuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gassnerland T., Gariboldi M., Giesi C., Godzik A., Gough J.,
RA Grimmond S., Gutinchich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lemhar B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Perlea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Walstedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
CC -!- FUNCTION: Cytoskeletal linker protein. Anchors keratin-containing
CC intermediate filaments to the inner plaque of hemidesmosomes. The
CC proteins may self-aggregate to form filaments or a two-dimensional
CC mesh (By similarity).
CC -!- SUBUNIT: Homodimer. Interacts with the neuronal intermediate
CC filament protein, Prph (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=7;
CC Name=2; Synonyms=b;
CC IsoId=Q91ZU6-1; Sequence=Displayed;
CC Name=1; Synonyms=a;
CC IsoId=Q91ZU6-2; Sequence=VSP_050483;
CC Name=3;
CC IsoId=Q91ZU6-3; Sequence=VSP_050484, VSP_050485, VSP_050486;
CC Note=No experimental confirmation available;
CC Name=4;
CC IsoId=Q91ZU6-4; Sequence=VSP_050485, VSP_050486;
CC Note=No experimental confirmation available;
CC Name=5; Synonyms=e;
CC IsoId=Q91ZU8-1; Sequence=External;
CC Name=6; Synonyms=nl;
CC IsoId=Q60824-1; Sequence=External;
CC Name=7; Synonyms=n2;
CC IsoId=Q60824-2; Sequence=External;
CC -!- TISSUE SPECIFICITY: Expressed at high levels in the heart and
CC skeletal muscle and at low levels in the skin in the adult.
CC Expressed in the myocardium, skeletal muscle masses, vertebrae
CC cartilage, and epithelia of the tongue of 14.5 day embryos.
CC -!- SIMILARITY: Belongs to the plakophilin or cytokeratin family.
CC -!- SIMILARITY: Contains 1 actin-binding domain.
CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC -!- SIMILARITY: Contains 9 plectin repeats.
CC -!- SIMILARITY: Contains 27 spectrin repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

[illegible]

|    |         |      |      |                         |
|----|---------|------|------|-------------------------|
| FT | REPEAT  | 6383 | 6480 | SPECTRIN 22.            |
| FT | REPEAT  | 6492 | 6592 | SPECTRIN 23.            |
| FT | REPEAT  | 6602 | 6701 | SPECTRIN 24.            |
| FT | REPEAT  | 6710 | 6808 | SPECTRIN 25.            |
| FT | REPEAT  | 6826 | 6914 | SPECTRIN 26.            |
| FT | REPEAT  | 6962 | 7020 | SPECTRIN 27.            |
| FT | CA BIND | 7028 | 7040 | EF-HAND 1 (POTENTIAL).  |
| FT | CA BIND | 7064 | 7076 | EF-HAND 2 (POTENTIAL).  |
| FT | DOMAIN  | 452  | 486  | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 638  | 703  | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 737  | 773  | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 1003 | 1138 | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 1195 | 1247 | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 1413 | 1455 | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 1504 | 1527 | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 3336 | 3359 | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 3539 | 3715 | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 3809 | 3893 | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 3957 | 3978 | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 4006 | 4043 | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 4159 | 4209 | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 4270 | 4311 | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 4426 | 4563 | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 4596 | 4735 | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 4847 | 5097 | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 5173 | 5233 | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 5570 | 5603 | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 5717 | 5739 | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 5787 | 5809 | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 6010 | 6089 | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 6116 | 6164 | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 6277 | 6318 | COILED COIL (POTENTIAL) |
| FT | DOMAIN  | 6381 | 6417 | COILED COIL (POTENTIAL) |

|                        |        |               |      |
|------------------------|--------|---------------|------|
| Pred. No.:             | 0.184  | Length:       | 7389 |
| Score:                 | 149.00 | Matches:      | 214  |
| Percent Similarity:    | 34.25% | Conservative: | 158  |
| Best Local Similarity: | 19.71% | Mismatches:   | 387  |
| Query Watch:           | 2.54%  | Indels:       | 327  |
| DB:                    | 1      | Gaps:         | 51   |

|     |  |      |   |      |
|-----|--|------|---|------|
| QY  |  | 170  | ATTTTTCTCCAGGATTCACGAAACCAAGATATCTCCATTACAATATACCATTCTGAAC    | 229  |
|     |  | :::  | :::   | :::  |
| Db  |  | 4839 | LeuLeuLeuSerThrGlnGlyAlaGluIysAlaIleuLeuGlnLeuAsnThrMet       | 4858 |
|     |  | :::  | :::   | :::  |
| QY  |  | 230  | AAAGACATAGACAATCCCCACAGAATGAACAACACTCAAAGTACTGAAAAAATG        | 283  |
|     |  |      | :::   | :::  |
| Ddb |  | 4859 | LysThrAspTrpAspArgPheArgLysGlnVallyGluArgGluGluIlysLeuIysAsp  | 4878 |
|     |  |      | :::   | :::  |
| QY  |  | 284  | -----TACAAAATGTCAACTCATATGAGACCAATATTCAT                      | 316  |
|     |  | :::  | :::   | :::  |
| Ddb |  | 4879 | SerLeuGluLySAlaLeuIlystyrArgGlnGlnValGluThrLeuArgProTrpileasp | 4998 |
|     |  | :::  | :::   | :::  |
| QY  |  | 317  | TTCGCCAAGCATACGAACAAAAGATCCGCATTTTTCCCACGGGGTGTTAAAGCTGTGCCA  | 376  |
|     |  | :::  |   | :::  |
| Ddb |  | 4899 | ArgCyGlnHisSerLeuaspGlyValThrPhe-----SerLeuaspPro             | 4913 |
|     |  | :::  |   | ---- |
| QY  |  | 377  | CAGGAATCCATCAAAACAGATT-----TTAGACAGTCTTCAA-----               | 412  |
|     |  |      |   | ---- |
| b   |  | 4914 | ThrGlusErGluSerSerilealaGluLeuIySerLeuGlnIySgUmetAspHisHis    | 4933 |
|     |  |      |   | ---- |
| QY  |  | 413  | -----GCTTATTATAGATTGAGATGTGTCTCAG-----                        | 439  |
|     |  |      |   | ---- |
| b   |  | 4934 | PheGlyMetLeuGluLeuLeuAsnAsnThrAlaAsnSerLeuLeuSerValCyGluVal   | 4953 |
|     |  |      |   | ---- |
| QY  |  | 440  | -----GAAGCAGTATGGGAGACATATCGAGTCTTTCTGGATCGCATCCCTGCACACAGG   | 493  |
|     |  |      |   | ---- |
| b   |  | 4954 | AspLysGluAlaValThrGluGluAsnGlnSerLeuMetGluIySValasArGValThr   | 4973 |
|     |  |      |   | ---- |
| QY  |  | 494  | GAATATCAGGACTGGGTGAGCATCTGCCAGGAGACCTTCTGCCTCTTTGACATTGGA     | 553  |





QY 2324 SCTGCAACACGATATGACAGCCAGGGAGCGCTCGACGGTCTGGAACAGCGCTCTGTG 2383  
 Db 5610 TGAAlaLysGluGluLeuGlu---LysThrIleAlaAspAspAenGluPro----- 5625  
 QY 2384 GCCCTGGCACAAGGAAGTGGAGTCTCCAGGGAAGGAGCTCCATGCGAGTGGCCAG 2443  
 Db 5626 -----LeuProAspCysGlu-----Prot 5632  
 QY 2444 ATCACTCTGAATCAACGATACAAACCTAGTGTGTAAAGATTCACAAATCAACAAATA 2503  
 Db 5632 hrGlnSerArgHisLysValGluGluIleAspAlaAlaIleLeuArgSerGlnGlnPhe 5652  
 QY 2504 ACAGGTATCATGTAAGAATCTGAATTAATCTGACCGTAGAATATGAGAATTAAC 2563  
 Db 5652 luGlnAla-----AlaAspAlaGluLeu----- 5659  
 QY 2564 ATCAAGATTGGGAAGAAATTAAGAACTGAAATGACAAATATCACTAGGCTATCTCA 2623  
 Db 5660 -----SerTrpIleThrGluThrGlnLysLysLeuMetSerLeuGlyAspIleArgLeu 5678  
 QY 2624 AGACAGATGATTGCCCTCTCAAGGAAATGGACAGCATATTCATGGTCTATC----- 2679  
 Db 5678 luGlnAspGlnThrSerAlaGlnLeuGlnValGlnLysAlaPheThrMetAspIleLeu 5698  
 QY 2680 -----AAAAATCCAGACATACAGTC 2698  
 Db 5698 rgHisLysAspIleAspGluLeuValThrSerGlyHisLysAlaMetThr-ThrSer 5717  
 QY 2699 AACACTGAGATCAACACACACATATTCATATATAGAGAGTCATG-----TACTTG 2752  
 Db 5718 GlyGluGluGlnLysGlnSerMetLysLysLysLysLysLysLysLysLysLysLys 5737  
 QY 2753 GCAACACGATAATCTGAAAGAAAGACACTTACTTATTTATTAAGCCCAATCAATC 2812  
 Db 5738 AlaValCysGlnLeuAsnSerGluArgHisLysGlnLeuGluArgAlaGlnSerLeuVal 5757  
 QY 2813 AGC-----GAAACATAT-----TTTACTATCTTCTGATGATA---GTCAATGATC 2857  
 Db 5758 SerGlnPheTrpGluThrTyrcGluLeuLeuProTrpLeuThrGluThrGlnArgile 5777  
 QY 2858 ATAGCCAGGTT 2869  
 Db 5778 IleSerGlnLeu 5781  
 RESULT 9  
 ID CRK7\_HUMAN STANDARD; PRT; 1490 AA.  
 AC Q9NYV4; O94978;  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Cell division cycle 2-related protein kinase 7 (SC 2.7.1.1-) (CDC2-  
 DE related protein kinase 7) (Crkrs).  
 GN CRK7 OR KIAA0904.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21539573; PubMed=11683387;  
 RA Ko T.K., Kelly E., Pines J.,  
 RT "Crkrs: a novel conserved Cdc2-related protein kinase that colocalises  
 RT with SC35 speckles";  
 RL J. Cell Sci. 114:2591-2603 (2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=22158633; PubMed=12168954;  
 RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;  
 RT "Construction of expression-ready cDNA clones for KIAA genes: manual  
 RT curation of 330 KIAA cDNA clones.";

RL DNA Res. 9:99-106 (2002).  
 RN [3]  
 RP PRELIMINARY SEQUENCE OF 266-1262 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=99156230; PubMed=10048485;  
 RA Nagase T., Iehikawa K.-I., Suyama M., Kikuno R., Hirose M.,  
 RA Miyajima N., Tanaka A., Kohani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XII.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RL for large proteins in vitro";  
 RL DNA Res. 5:355-364 (1998).  
 CC -|- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -|- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.  
 CC CDC2/CdkX subfamily.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; AF227198; AAF36401.1;  
 DR EMBL; AB020711; BAA74927.2; ALT\_INIT.  
 DR HSSP; P24941; IBOH.  
 DR GK; Q9NYV4; -;  
 DR InterPro; IPR000719; Prot\_kinase.  
 DR InterPro; IPR008271; Ser\_Thr\_kinase.  
 DR InterPro; IPR002290; Ser\_Thr\_kinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR SMART; SM00220; S\_TK; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
 DR PROSITE; PS50011; PROTEIN KINASE\_DOM; 1.  
 KW Transferase; Serine/threonine-protein kinase; ATP-binding;  
 KW Nuclear protein.  
 FT DOMAIN 727 1020 PROTEIN KINASE.  
 FT NP\_BIND 733 741 ATP (BY SIMILARITY).  
 FT BINDING 756 756 ATP (BY SIMILARITY).  
 FT ACT\_SITE 859 859 BY SIMILARITY.  
 FT DOMAIN 407 413 POLY-ALA.  
 FT DOMAIN 535 540 POLY-PRO.  
 FT DOMAIN 1266 1280 POLY-PRO.  
 FT CONFLICT 639 639 R -> D (IN REF. 2 AND 3).  
 FT CONFLICT 745 745 R -> K (IN REF. 2 AND 3).  
 FT CONFLICT 1254 1262 MISSING (IN REF. 2).  
 SQ SEQUENCE 1490 AA; 164154 MW; 851E18DF3BD2B1A1 CRC64;  
 Alignment Scores:  
 Pred. No.: 0.132 Length: 1490  
 Score: 148.00 Matches: 119  
 Percent Similarity: 33.70% Conservative: 63  
 Best Local Similarity: 22.04% Mismatches: 177  
 Query Match: 2.52% Indels: 182  
 DB: 1 Gaps: 23  
 US-10-007-270-1 (1-3330) x CRK7\_HUMAN (1-1490)  
 QY 427 GAGAGTGTGTGAGGAGCAGTATGGGAACATATCGATCTTCTGGATGCGA-TCCTG 485  
 Db 235 AspSerProSerGlyAlaSerTyrcGlnAspTyrcAspLeuSerProSerArgSerHis 254  
 QY 486 ACACAGGGGAAT-----ATCAGACTGGTCCAGCTCCAGCAGGAGACCTCT 536  
 Db 255 ThrSerSerAsnTyrcAspSerTyrcLysLysSerProGlySerThrSerArgGlnSer 274  
 QY 537 GCCTCTTTCATCTGGAAAAAATCTCAGCAATCCAGGAGCAGCCTGATCTTCTCCAG 596  
 Db 275 ValSerProProTyrcLysGluProSerAlaTyrcGlnSerSerThr----- 289  
 QY 597 AGAGAATAAACAAGAGAGTTTCCCTGACAGAAAGATGAATATCTCGCAGAGACAT 656

Db 290 -----ArgSerProSerProTyArgSer----- 298  
 QY 657 TGGAGAGCTGGTGAACACCATTTCTTCAACAGCAATCTACATTTCAAGACTGGG 716  
 Db 299 ---GlnArgSerValSerProTyArgSerProTyArgSerSerTyGluArgSerGly 317  
 QY 717 CAGTAT----- 722  
 Db 318 SerTySerGlyArgSerProSerProTyArgSerSerSerProPheLeu 337  
 QY 723 TCTAAGAAACCTCTCAGAGAGCA-----AATCAAGATGTTGCCACGCTCTCACTGG 776  
 Db 338 SerTyArgSerLeuSerArgSerProLeuProSerArgLysSerMetLysSerArgSer 357  
 QY 777 GCCTTCCTCTCTCACTCTGATGACACCTCTCTCAATGAAATTCCTCGATATACACTCAA 836  
 Db 358 ArgSerProAlaTySer----- 363  
 QY 837 CGACCAAGATGCTTACACAGAAAG-----AGAAACAGAAATTCCTCT 881  
 Db 364 ArgHisSerSerHisSerLysLysLysArgSerSerArgSerArg-HisSerSe 383  
 QY 882 GTTGGAGGACAGAGGGTGGAGCTCAGCGCTCTCTCTGTAACACAGAGTTCAGGCAGA 941  
 Db 383 rIleSerProValArgLeuProLeuAsnSerLeu-----GlyAlaG 398  
 QY 942 GCTCGCTGATCCCAAGTCCCATATTACAGGAGTAGCAGGAAGTCCCACTTCAGAT 1001  
 Db 398 uLeuSerArgLys-----LysLysGluArgAla 408  
 QY 1002 GCAAGAGATATTAAAGAACTCCAGGATTCAAAATAATCCATGGTTAGATTAGACC 1061  
 Db 408 alaAlaalaalalaLysMetAspGlyLysGluSerLysGlySerProValPheLeuPr 428  
 QY 1062 AAAGAAAGAAAGATGGCTCAAGCTCCACAGAGATGCAACTTACGGCCATCTTAAGG 1121  
 Db 428 oArgLysGluAsnSer----- 433  
 QY 1122 ACACAGTCAAGCAACCAAGCCCTGCAAGTGACCTCTGTCTTTGATTCCAAACAAAT 1181  
 Db 434 ---SerValGluAlaLys-----AspSerGly-----Le 442  
 QY 1182 TGAAGTGAAGAGTCTATCATGGAACCTGAGGAGGAGCAAGCAACAGAAATCTATCT 1241  
 Db 442 uGluSerLysLysLeuProArgSerValLysLeuGluLysSerAlaProAsp----- 459  
 QY 1242 CACAGCTACAGACTCAAAAGCTGATCAGCAAGCACTAGAGGAGCAACATCTTTGGA 1301  
 Db 460 -----ThrGluLeuValAsnValThrHisLeuAsnThrGluValLysAsnSerSerAs 477  
 QY 1302 TGTGGGACAAATTCAGTCTCAGTAT-----GAAATTCGTGATCCTGCCAGC 1349  
 Db 477 pThrGlyValLysValLysLeuAspGluAsnSerGluLysHisLeuValLysAspLeuLysAl 497  
 QY 1350 CTTTGGTCTCTGACACCAATCAGAGCTGCCACATCTTTCTGTTATACA-----GA 1403  
 Db 497 agLysGlyThrArgAspSerLysProIleAlaLeuLysGluGluValThrProLysG 517  
 QY 1404 GATGCTACTTGGTCCAGCACTTCTCTGTTGAACCCAGCTTGAACAGTGGACGG 1463  
 Db 517 uThrGluThrSerGluLysGluThrPro-----ProLeuProThrIleAlaSe 534  
 QY 1464 ACAGAGATGGTCTACTCTGACACTTCTTGCT-----CCACTGC 1505  
 Db 534 rProProProProLeuProToThrThrThrProProProGlnThrProProLeuProLe 554  
 QY 1506 TATGGCTCTACTCTCTGTCAGAGCTCCACTTCTTTATGGCATCAAGCTCTCTC 1565  
 Db 554 uProProIleProAlaLeuProGlnThrProProProSerGlnProAlaPheSe 574  
 QY 1566 TCTGACTGATCAAGCACCACAGATCAATGGCCACTGACCAAGATGCTAGTACCAGG 1625

Db 574 rGlnValProAlaSerSerThrSer----- 582  
 QY 1626 GCTCACCATCCCAACAGATGATTTCTGCAATCAGCAACTG---GCTCTGGGAATTC 1682  
 Db 583 ---ThrLeuProProSerThrHisSerLysThrSerAlaValSerSerGlnAlaAsnSe 601  
 QY 1683 ACATCCCACT-----GCATCTTCAGATGA 1706  
 Db 601 rGlnProProValGlnValSerValLysThrGlnValSerValThrAlaAlaIleProHi 621  
 QY 1707 CAGCCGATCAAGTGCA-----GGTGGCGAGA 1733  
 Db 621 sLeuLysThrSerThrLeuProLeuProLeuProLeuProGlyGlyAspAs 641  
 QY 1734 TATG----- 1737  
 Db 641 pMetAspSerProLysGluThrLeuProSerLysProValLysGluLysGluGlnAr 661  
 QY 1738 -GTCAAGACACCTAGATGAATGGATCTGTCTGACACTCTGCCCATCTGAGGTACCA 1794  
 Db 661 gThrArgHisLeu-----LeuThrAspLeuProLeuProGluLeuPro 676  
 RESULT 10  
 ID PCGV RAT STANDARD; PRT; 2738 AA.  
 AC Q9ERB4; O08592; O88564; Q9R1K4;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Versican core protein precursor (Large fibroblast proteoglycan)  
 DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial  
 DE hyaluronate-binding protein) (GHAAP) (Fragments).  
 GN CP5G2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0), SEQUENCE FROM N.A.  
 RP (ISOFORM V3), AND SEQUENCE OF 2657-2738 FROM N.A. (ISOFORM VINT).  
 RC STRAIN=Wistar Kyoto;  
 RX MEDLINE=99327053; PubMed=10397680;  
 RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,  
 RA Wight T.N.;  
 RT "Versican/PG-M isoforms in vascular smooth muscle cells.";  
 RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).  
 RN [2]  
 RP SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0).  
 RC STRAIN=Wistar Kyoto;  
 RX MEDLINE=98308094; PubMed=9642104;  
 RA Milev P., Maurel P., Chiba A., Mevissen M., Popp S., Yamaguchi Y.,  
 RA Margolis R.K., Margolis R.U.;  
 RT "Differential regulation of expression of hyaluronan-binding  
 RT proteoglycans in developing brain: aggrecan, versican, neurocan, and  
 RT brevican.";  
 RL Biochem. Biophys. Res. Commun. 247:207-212(1998).  
 RN [3]  
 RP SEQUENCE OF 2421-2463 FROM N.A. (ISOFORM V0).  
 RC TISSUE=Kidney;  
 RX MEDLINE=98094159; PubMed=9434070;  
 RA Pyke C., Kristensen P., Ostergaard P.B., Oturai P.S., Romer J.;  
 RT "Proteoglycan expression in the normal rat kidney.";  
 RL Nephron 77:461-470(1997).  
 RN [4]  
 RP SEQUENCE OF 2535-2738 FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Lung;  
 RA Blomberg L.A., Chan W.-Y., Clerch L., Massaro D.;  
 RT "Molecular cloning and characterization of two developmentally  
 RT regulated genes in rat lung.";  
 RL Submitted (Sep-2000) to the EMBL/GenBank/DBJ databases.  
 CC -i- FUNCTION: May play a role in intercellular signaling and in  
 CC connecting cells with the extracellular matrix. May take part in  
 CC the regulation of cell motility, growth and differentiation. Binds

**hyaluronic acid.**

-|- SUBUNIT: Interacts with PELNI (By similarity).

-|- SUBCELLULAR LOCATION: Secreted; extracellular matrix.

-|- ALTERNATIVE PRODUCTS:

Event-Alternative splicing; Named isoforms=3;

Comment=Additional isoforms seem to exist;

Name=V0;

IsoId=Q9ERB4-1; Sequence=Displayed;

Name=V3;

IsoId=Q9ERB4-2; Sequence=VSP\_003091;

Name=Vint;

IsoId=Q9ERB4-3; Sequence=VSP\_003092;

-|- TISSUE SPECIFICITY: In kidney is expressed in the papillary area,  
but not in glomeruli.

-|- DEVELOPMENTAL STAGE: Disappears after the cartilage development  
(By similarity).

-|- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.

-|- SIMILARITY: Contains 2 link domains.

-|- SIMILARITY: Contains 2 EGF-like domains.

-|- SIMILARITY: Contains 1 C-type lectin family domain.

-|- SIMILARITY: Contains 1 Sushi (SCR) domain.

-|- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/annouce/ or send an email to license@isb-sib.ch).

EMBL; AF062402; AAC40166.1; -

EMBL; U75306; AAB51125.1; -

EMBL; AF084544; AAD48544.1; -

EMBL; AY072892; AAC26116.1; -

EMBL; AY007691; AAG16631.1; -

HSSP: P01132; 1EPG.

InterPro: IPR000152; Asx hydroxyl\_s.

InterPro: IPR000742; EGF 2.

InterPro: IPR001881; EGF Ca.

InterPro: IPR006209; EGF\_like.

InterPro: IPR007110; Ig-Like.

InterPro: IPR003599; Ig.

InterPro: IPR001304; Lectin C.

InterPro: IPR000538; Link.

PRINTS: PR01265; LINKMODULE.

ProDom: PD000918; Link; 2.

SMART; SM00032; CCP; 1.

SMART; SM00034; CLECT; 1.

SMART; SM00179; EGF\_CA; 1.

SMART; SM00409; IG; 1.

SMART; SM00445; LINK; 2.

PROSITE; PS00010; ASX HYDROXYL; 1.

PROSITE; PS00615; C TYPE LECTIN 1; 1.

PROSITE; PSS0041; C TYPE LECTIN 2; 1.

PROSITE; PS00022; EGF 1; 2.

PROSITE; PS01186; EGF 2; 1.

PROSITE; PS00026; EGF 3; 2.

PROSITE; PS01187; EGF CA; 1.

PROSITE; PSS0835; IG LIKE; 1.

PROSITE; PS01241; LINK; 2.

Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;  
Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;  
Hyaluronic acid; Alternative splicing.  
SIGNAL 1 20 POTENTIAL.

CHAIN 21 2738

FT NON CONS 348 349

FT DOMAIN 21 146

FT DOMAIN 167 244

FT DOMAIN 265 346

FT DOMAIN <349 695

GAG-ALPHA  
(GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN).



|    |      |  |      |
|----|------|--|------|
| QY | 2012 | AGTGTCTGTTCCATATCATCATCAATCTTACAGGATTAAAGCAACTTGAATACTTA       | 2071 |
| Db | 2377 | lnArgileLeuAspTyr-----SerAsnGlnAlaThrValSerThrLeuAsp-----LeuA  | 2394 |
| QY | 2072 | ACTTCAGAACGGG-----AGTGTGATGTGGAATAGCAAAATCAAGATTGTCTG          | 2119 |
| Db | 2394 | snThrGluHisSerileProProPheSerileLeuAspAsnSer-----              | 2408 |
| QY | 2120 | AGTCTGTGCGGTATAAAGCTCACCAGAGGTGTGCACGGGCTCTGGAGGATTTTCGTTCTG   | 2179 |
| Db | 2409 | -----AsnGluThrAlaPheLeuLeuGlyIleSerGluGluThrValGluG            | 2424 |
| QY | 2180 | CTCAGCCCAACAACCTCCACTCTGGAATATGACAGCTACTCTCTCACATTGAACACGCTG   | 2239 |
| Db | 2424 | lyThrAlaValTyrLeu-----   | 2430 |
| QY | 2240 | ATCACAGCATCTCCGACAGTTCCTGGCTGCGCGGCAATTTGCCCAATGTGTAAGAACG     | 2299 |
| Db | 2430 | roGlyProAspLeuCySylsThrAsnProCySylLeuAsnGlyGlyThrCysTyrProThrG | 2450 |
| QY | 2300 | AACGACTCAGGAGCGAGTGTCTGCTGCAAAACCCAGATAT-----                  | 2340 |
| Db | 2450 | luThrSer-----TyrValCysThrCysAlaProGlyTyrSerGlyAspGlnCysGluL    | 2468 |
| QY | 2341 | -----GACAGCCAGGAGGACCTCGACACGCTC                               | 2365 |
| Db | 2468 | euAspPheAspGluCySHisSerAsnProCyArgAsnGlyAlaThrCysValAspGlyL    | 2488 |
| QY | 2366 | TGGAACCA-----GGCTCTGTGGCCCTCGCCA                               | 2392 |
| Db | 2488 | euAsnThrPheArgCysLeuProSerTyrValGlyAlaLeuCySgluGlnAspT         | 2508 |
| QY | 2393 | CAAGGAATCGAGTCT-----CTCCAGGGAAG-----GGAGCTC                    | 2428 |
| Db | 2508 | hrGluThrCysAspTyrGlyTrpHisIlePheGlnGlyCysTyrIleTyrPheAlaH      | 2528 |
| QY | 2429 | CATGCAAGTTCGACATCATCTCTGAAATCAA-----GCATACAAACTA               | 2473 |
| Db | 2528 | isArgArgThrTrpAspAlaIleGluArgGluCysArgLeuGlnGlyAlaHisLeuThrS   | 2548 |
| QY | 2474 | GTGTATAAAGTTCCAAATCAACA-----                                   | 2499 |
| Db | 2548 | erIleLeuSerHisGluGluGlnMetPheValAsnArgValGlyHisAspTyrGlnTriP   | 2568 |
| QY | 2500 | -----ATAACAGGTATCAGTAAA-----AGAAATTCGAATTACTGACCGTAG           | 2545 |
| Db | 2568 | leglyLeuAsnAspLysMetPheGluHisAspPheArgTrpThrAspGlySerAlaLeuG   | 2588 |
| QY | 2546 | AATATGAGAAATT-----AACATCATAGAT-----                            | 2571 |
| Db | 2588 | lnTyrGluAsnTrpArgProAsnGlnProAspSerPhePheSerAlaGlyLeuAspCysV   | 2608 |
| QY | 2572 | -----TGGGAAGGAAATTAATAACTGAAA-----ATGTACATATTATCACTTAG         | 2614 |
| Db | 2608 | alValIleIleTrpHisGluAsnGlyGlnTrpAsnAspValProCysAsnTyrHisLeuT   | 2628 |
| QY | 2615 | GCTAT---CTCAAGACAGATGATTTG-----CCTTCTCAGGGAATATGGA             | 2656 |
| Db | 2628 | hrTyrThrCysIleLysGlyThrValAlaCysGlyGlnProProValValGluAsnAla    | 2647 |
| QY | 2657 | GACAGCATATTCAGGTCATCAAAATCCAGACATACATGACACTGAGATATCAGCAC       | 2716 |
| Db | 2648 | -----LysThrPheGlyLysMetLysProArgTyrGluIleAsnSer-----           | 2661 |
| QY | 2717 | ACACCATATTCATAGACAGTCATGCTTGGCAACCACTGAATAATCTGAAAAAAA         | 2776 |
| Db | 2662 | -----LeuIleArgTyrHisCysLysAspGlyPhe---IleGln                   | 2673 |
| QY | 2777 | AGACCTACTTATTATAA-----   | 2797 |
| Db | 2674 | ArgHisLeuProThrIleArgCysLeuGlyAsnGlyArgTrpAlaMetProLysIleThr   | 2693 |

2798 -----ACCCCAATGCAATCAGCGAAACATAT 2824  
||||:||||  
2694 CysMetAanProSerAlaTyr-GlnArgThrTyr 2704  
|||||

RESULT 11

| DB | AMNH_YEAST   | STANDARD; | PRT; 1367 AA.      |
|----|--|-----------|--------------------|
| AC | P08640; P08068;  |           |                    |
| DT | 01-AUG-1988 (Rel. 08, Created)   |           |                    |
| DT | 01-FEB-1995 (Rel. 31, Last sequence update)  |           |                    |
| DT | 10-OCT-2003 (Rel. 42, Last annotation update)  |           |                    |
| DE | Glucosylase S1/S2 precursor (EC 3.2.1.3) (Glucon 1,4-alpha-glucosidase) (1,4-alpha-D-glucan (glucohydrolase)).   |           |                    |
| DE | STAI OR STA2 OR MAL5 OR YI019C.  |           |                    |
| GN | Saccharomyces cerevisiae (Baker's yeast).  |           |                    |
| OC | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;   |           |                    |
| OC | Saccharomycetales; Saccharomycetaceae; Saccharomyces.  |           |                    |
| OX | NCBI_TaxID=4932;   |           |                    |
| EN | [1]  |           |                    |
| RP | SEQUENCE FROM N.A.   |           |                    |
| RC | STRAIN=S288c / AB972;  |           |                    |
| FX | MEDLINE=97313266; PubMed=9169870;  |           |                    |
| RA | Churcher C.M., Bowman S.R., Badcock K., Bankier A., Brown D.,  |           |                    |
| RA | Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,   |           |                    |
| RA | Harris D.B., Horsnell T., Hunt S., Jagels K., Jones M., Lye G.,  |           |                    |
| RA | Moule S., Odeh C., Pearson D., Rajandream M.A., Rice P., Rowley N.,  |           |                    |
| RA | Skelton J., Smith V., Walsh S., Whitehead S., Barrett B.G.,  |           |                    |
| RT | "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX."   |           |                    |
| RT | Nature 387:84-87(1997).  |           |                    |
| RL | [2]  |           |                    |
| RP | SEQUENCE OF 1-242 AND 762-1331 FROM N.A.   |           |                    |
| RC | MEDLINE=97194600; PubMed=3106330;  |           |                    |
| RA | Yamashita I., Nakamura M., Fukui S.;   |           |                    |
| RT | "Gene fusion is a possible mechanism underlying the evolution of   |           |                    |
| RT | STAL".   |           |                    |
| RL | J. Bacteriol. 169:2142-2149(1987).   |           |                    |
| RP | [3]  |           |                    |
| RP | SEQUENCE OF 1-31 FROM N.A.   |           |                    |
| RC | STRAIN=SPX101-1C;  |           |                    |
| RA | MEDLINE=89031230; PubMed=3142113;  |           |                    |
| RA | Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;  |           |                    |
| RT | "Similar short elements in the 5' regions of the STA2 and SGA genes  |           |                    |
| RT | from Saccharomyces cerevisiae".  |           |                    |
| RL | FEBS Lett. 239:179-184(1988).  |           |                    |
| CC | -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-   |           |                    |
| CC | glucose residues successively from non-reducing ends of the chain  |           |                    |
| CC | with release of beta-D-glucose.  |           |                    |
| CC | -!- SIMILARITY: TO S.POMBE SPBC215.13.   |           |                    |
| CC | -!- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.  |           |                    |
| CC |  |           |                    |
| CC | This SWISS-PROT entry is copyright. It is produced through a collabora-  |           |                    |
| CC | tion between the Swiss Institute of Bioinformatics and the EMBL outstation                                       |           |                    |
| CC | at the European Bioinformatics Institute. There are no restrictions on   |           |                    |
| CC | use by non-profit institutions as long as its content is in no   |           |                    |
| CC | modified and this statement is not removed. Usage by and for commer-   |           |                    |
| CC | entities requires a license agreement (See <a href="http://www.isb-sib.ch/annou">http://www.isb-sib.ch/annou</a> |           |                    |
| CC | or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).                                |           |                    |
| CC |  |           |                    |
| DR | EMBL; Z38061; CAA86176.1; -  |           |                    |
| DR | EMBL; M16164; AAA35014.1; -  |           |                    |
| DR | EMBL; M16165; AAA35015.1; -  |           |                    |
| DR | EMBL; X13857; CAA32069.1; -  |           |                    |
| DR | PIR; S48478; S48478.   |           |                    |
| DR | GermOnline; L39731; -  |           |                    |
| DR | SGD; S0001458; MUC1.   |           |                    |
| DR | GO; GO:0005886; C:plasma membrane; IDA.  |           |                    |
| DR | GO; GO:0030447; P:filamentous growth; IDA.   |           |                    |
| DR | GO; GO:0007125; P:invasive growth; IMP.  |           |                    |
| DR | GO; GO:0007124; P:pseudohyphal growth; IMP.  |           |                    |
| KW | Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;  |           |                    |
| CC | Signal; Multigene family.  |           |                    |
| CC | SIGNAL   |           |                    |
| CC | 1 21   |           | POTENTIAL,         |
| CC | 22 1367  |           | GLUCOSYLASE S1/S2. |
| CC | CHAIN  |           |                    |

|    |      |   |      |
|----|------|---|------|
| Db | 555  | roThrProSerSer-----SerThrThrGluSerSerSerThrProValThrSerSert   | 573  |
| Qy | 1601 | CTGACCAGACAATGCTAGTACCAGGGCTCCACCATCCACCAGGATATTCTGCAATCA     | 1660 |
| Db | 573  | hrThrGluSerSerSerAlaPro-----ValProThrProSerSerSerThrThrG      | 590  |
| Qy | 1661 | GCCAACTGGCTCTGGGAATTCATCCACTGCGATCTTCAGATGACAGCCGATCAATG      | 1720 |
| Db | 590  | luSerSerSerAlaProValProThrProSerSerSerThrThrGluSer-----       | 606  |
| Qy | 1721 | CAGGTGGGAAGATATGTCAGACACCTAGATGAATGGATCTGTCTGACACTCTCGCC      | 1780 |
| Db | 607  | -----SerSerAlaProAlaP   | 612  |
| Qy | 1781 | CATCTGAGGTACCAGAGCTCAGCGAATATGTTCTGTCCTCCAGATCATTTCTTGGAGGATA | 1840 |
| Db | 612  | toThrProSerSerThrThrGluSerSerSerAlaPro-----ValThrSerSert      | 630  |
| Qy | 1841 | CCACT-----CCTGCTCAGCTTACAGTATATCAACACT-----AGTT               | 1879 |
| Db | 630  | hrThrGluSerSerSerAlaProValProThrProSerSerSerThrThrGluSerSerS  | 650  |
| Qy | 1880 | CTATGACCAATGCCC-----CCAAGGGCGAGAGCTGGTAGTGTCTTCAGTCTGCGTGTG   | 1936 |
| Db | 650  | erAla-ProValProThrProSerSerSerThrThrGluSerSerSerAlaProVal---  | 668  |
| Qy | 1937 | CTAAATGCGCTTCCAAAGCACTCTTCAACAGAGCTCTCTGGAGTACCGAGCTCTGG      | 1996 |
| Db | 669  | -----ProThrProSerSerSerThrThrGluSerSerSerAlaProValThrSe       | 685  |
| Qy | 1997 | AGCAACAATTCACAGCTGCTGGTCCATA-----TTCAGATCCAACT                | 2042 |
| Db | 685  | rSerThrThrGluSerSerSerAlaProValThrSerSerSerThrThrGluSer       | 702  |

RESULT 12

PGCV HUMAN

ID PGCV HUMAN STANDARD; PRT; 3396 AA.

AC P13611; P20754; Q13010; Q13189; Q15123; Q9UNW5;

DT 01-NOV-1990 (Rel. 13, Created)

DT 01-JAN-1997 (Rel. 35, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Versican core protein precursor (large fibroblast proteoglycan)

DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial

DE hyaluronate-binding protein) (GHA).

GN CSPG2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OC NCBI\_TaxID=9606;

RP [1]

RP SEQUENCE FROM N.A. (ISOFORM V0).

RX MEDLINE=95105188; PubMed=7528742;

RX Nabo M.F., Zimmermann D.R., Iozzo R.V.;

RT "Characterization of the complete genomic structure of the human

RT versican gene and functional analysis of its promoter.";

RL J. Biol. Chem. 269:32999-33008(1994).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORM V1).

RX TISSUE=Placenta;

RX MEDLINE=90059882; PubMed=2583089;

RX Zimmermann D.R., Ruoslahti E.;

RT "Multiple domains of the large fibroblast proteoglycan, versican.";

RL EMBO J. 8:2975-2981(1989).

RN [3]

RP SEQUENCE FROM N.A. (ISOFORM V2).

RX TISSUE=Glial tumor;

RX MEDLINE=95105187; PubMed=7806529;

RX Dours-Zimmermann M.T., Zimmermann D.R.;

RT "A novel glycosaminoglycan attachment domain identified in two

RT alternative splice variants of human versican.";

RL J. Biol. Chem. 269:32992-32998(1994).

RN [4]

RP SEQUENCE OF 2711-3396 FROM N.A.

RC TISSUE=Lung fibroblast;  
 RX MEDLINE=88007514; PubMed=2820964;  
 RA Krusius T., Gehlsen K.R., Ruoslahti E.;  
 RT "A fibroblast chondroitin sulfate proteoglycan core protein contains  
 RT lectin-like and growth factor-like sequences."; [5]  
 RL J. Biol. Chem. 262:13120-13125(1987).  
 RN [5]  
 RC SEQUENCE OF 251-347 FROM N.A.  
 RX MEDLINE=93122792; PubMed=1478664;  
 RA Lozzo R.V., Naso M.F., Cannizzaro L.A., Wasmuth J.J.,  
 RA McPherson J.D.;  
 RT "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of  
 RT human chromosome 5 (5q12-5q14)."; [6]  
 RL Genomics 14:845-851(1992).  
 RN [6]  
 RC SEQUENCE FROM N.A. (ISOFORM V3).  
 RX TISSUE=Brain;  
 RA MEDLINE=95181355; PubMed=7876137;  
 RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;  
 RT "Expression of PG-M(V3), an alternatively spliced form of PG-M  
 RT without a chondroitin sulfate attachment in region in mouse and human  
 RT tissues."; [7]  
 RL J. Biol. Chem. 270:3914-3918(1995).  
 RN [7]  
 RC SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).  
 RX TISSUE=Aortic smooth muscle;  
 RA MEDLINE=99327053; PubMed=10397680;  
 RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,  
 RA Wright T.N.;  
 RT "Versican/Pg-M isoforms in vascular smooth muscle cells."; [8]  
 RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).  
 RN [8]  
 RC PARTIAL SEQUENCE.  
 RX TISSUE=Brain;  
 RA MEDLINE=89174663; PubMed=2466833;  
 RA Perides G., Lane W.S., Andrews D., Dahl D., Bignami A.;  
 RT "Isolation and partial characterization of a glial  
 RT hyaluronate-binding protein."; [9]  
 RL J. Biol. Chem. 264:5981-5987(1989).  
 RN [9]  
 RC TISSUE SPECIFICITY OF ISOFORMS.  
 RX MEDLINE=96213482; PubMed=8627343;  
 RA Paulus W., Baur I., Dours-Zimmermann M.T., Zimmermann D.R.;  
 RA "Differential expression of versican isoforms in brain tumors.";  
 RT J. Neuropathol. Exp. Neurol. 55:528-533(1996).  
 CC -I- FUNCTION: May play a role in intercellular signaling and in  
 CC connecting cells with the extracellular matrix. May take part in  
 CC the regulation of cell motility, growth and differentiation. Binds  
 CC hyaluronic acid.  
 CC -I- SUBUNIT: Interacts with FBLN1 (By similarity).  
 CC -I- SUBCELLULAR LOCATION: Secreted; extracellular matrix.  
 CC -I- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=5;  
 CC Comment=Additional isoforms seem to exist;  
 CC Name=V0;  
 CC IsoId=P13611-1; Sequence=Displayed;  
 CC Name=V1;  
 CC IsoId=P13611-2; Sequence=VSP\_003082, VSP\_003083;  
 CC Name=V2;  
 CC IsoId=P13611-3; Sequence=VSP\_003084;  
 CC Name=V3;  
 CC IsoId=P13611-4; Sequence=VSP\_003082, VSP\_003085;  
 CC Name=Vint;  
 CC IsoId=P13611-5; Sequence=VSP\_003086;  
 CC -I- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 is expressed  
 CC in normal brain, gliomas, medulloblastomas, schwannomas,  
 CC neurofibromas, and meningiomas; v2 is restricted to normal brain  
 CC and gliomas; v3 is found in all these tissues except  
 CC medulloblastomas.  
 CC -I- DEVELOPMENTAL STAGE: Disappears after the cartilage development.  
 CC -I- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
 CC -I- SIMILARITY: Contains 2 link domains.  
 CC -I- SIMILARITY: Contains 2 EGF-like domains.

CC -I- SIMILARITY: Contains 1 C-type lectin family domain.  
 CC -I- SIMILARITY: Contains 1 Sushi (SCR) domain.  
 CC -I- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U16306; AAA65018.1; -  
 CC EMBL: X15998; CAA34128.1; -  
 CC EMBL: S52488; AAB24878.1; -  
 CC EMBL: U26555; AAA67565.1; -  
 CC EMBL: D32039; EAA06801.1; -  
 CC EMBL: J02814; AAA36437.1; -  
 CC EMBL: AF084545; AAD48545.1; -  
 CC PIR: S06014; A60979.  
 CC HSSP: P01132; IEGF.  
 CC Genew: HGNC:2464; CSPG2.  
 CC MIM: 118661; -  
 CC GO: GO:0005578; C:extracellular matrix; TAS.  
 CC GO: GO:0005540; F:hyaluronic acid binding; TAS.  
 CC GO: GO:0008037; P:cell recognition; TAS.  
 CC GO: GO:0007275; P:development; TAS.  
 CC InterPro: IPR000152; Asx\_hydroxyl\_S.  
 CC InterPro: IPR000742; EGF 2.  
 CC InterPro: IPR001881; EGF\_Ca.  
 CC InterPro: IPR006209; EGF\_like.  
 CC InterPro: IPR007110; Ig\_Like.  
 CC InterPro: IPR003599; Ig.  
 CC InterPro: IPR001304; Lectin\_C.  
 CC InterPro: IPR000538; Link.  
 CC InterPro: IPR000436; Sushi\_SCR\_CCP.  
 CC Pfam: PF00008; EGF; 2.  
 CC Pfam: PF00059; lectin\_C; 1.  
 CC Pfam: PF00084; sushi; 1.  
 CC Pfam: PF00193; Link; 2.  
 CC PRINTS: P01265; LINKMODULE.  
 CC ProDom: PD000918; Link; 2.  
 CC SMART: SM00032; CCP; 1.  
 CC SMART: SM00034; CLECT; 1.  
 CC SMART: SM00179; EGF\_CA; 1.  
 CC SMART: SM00409; IG; 1.  
 CC SMART: SM00445; LINK; 2.  
 CC PROSITE: PS00010; ASX\_HYDROXYL; 1.  
 CC PROSITE: PS00615; C\_TYPE\_LECTIN\_1; 1.  
 CC PROSITE: PS00041; C\_TYPE\_LECTIN\_2; 1.  
 CC PROSITE: PS00022; EGF\_1; 2.  
 CC PROSITE: PS01186; EGF\_2; 1.  
 CC PROSITE: PS00026; EGF\_3; 2.  
 CC PROSITE: PS01187; EGF\_CA; 1.  
 CC PROSITE: PS00835; IG\_Like; 1.  
 CC PROSITE: PS01241; LINK; 2.  
 CC Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;  
 CC Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;  
 CC Hyaluronic acid; Alternative splicing.  
 CC SIGNAL 1 20  
 CC POTENTIAL.  
 CC FT CHAIN 21 3396 VERSICAN CORE PROTEIN.  
 CC FT DOMAIN 21 146 IG-LIKE V-TYPE.  
 CC FT DOMAIN 167 244 LINK 1.  
 CC FT DOMAIN 265 346 LINK 2.  
 CC FT DOMAIN 348 1335 GAG-ALPHA  
 CC (GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN).  
 CC FT DOMAIN 1336 3089 GAG-BETA.  
 CC FT DOMAIN 3089 3125 EGF-LIKE 1.  
 CC FT DOMAIN 3127 3163 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
 CC FT DOMAIN 3176 3290 C-TYPE LECTIN.  
 CC FT DOMAIN 3295 3353 SUSHI.  
 CC FT DISULFID 44 130 BY SIMILARITY.  
 CC FT DISULFID 172 243 BY SIMILARITY.

FT DISULFID 196 217 BY SIMILARITY.  
 FT DISULFID 270 345 BY SIMILARITY.  
 FT DISULFID 294 315 BY SIMILARITY.  
 FT DISULFID 3093 3104 BY SIMILARITY.  
 FT DISULFID 3098 3113 BY SIMILARITY.  
 FT DISULFID 3115 3124 BY SIMILARITY.  
 FT DISULFID 3131 3142 BY SIMILARITY.  
 FT DISULFID 3136 3151 BY SIMILARITY.  
 FT DISULFID 3153 3162 BY SIMILARITY.  
 FT DISULFID 3169 3180 BY SIMILARITY.  
 FT DISULFID 3197 3289 BY SIMILARITY.  
 FT DISULFID 3265 3281 BY SIMILARITY.  
 FT DISULFID 3296 3339 BY SIMILARITY.  
 FT DISULFID 3325 3352 BY SIMILARITY.  
 FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).

Alignment Scores:  
 Pred. No.: 0.209 Length: 3396  
 Score: 146.50 Matches: 184  
 Percent Similarity: 29.66% Conservative: 115  
 Best Local Similarity: 18.25% Mismatches: 313  
 Query Match: 2.49% Indels: 396  
 DB: 1 Gaps: 48

US-10-007-270-1 (1-3330) x PGCV\_HUMAN (1-3396)

QY 691 AGCAATCTACATTTCAAAGACTTGGGCGAGTATCTCAAGAAACCTCAGAAGACCAATT 750  
 Db 2455 SerSerThrThrPheValSerAspGlySerLeuGluLysHisPro-----GluVal 2471  
 QY 751 CRAAGTGTGGCAACGCTCACTTGGGCTTTTCCTCTCACT----- 792  
 Db 2472 ProSerAlaLysAlaValThrAlaAspGlyPheProThrValSerValMetLeuProLeu 2491  
 QY 793 -----CCTGATGACACCCCTCCCTCAAT----- 813  
 Db 2492 HisSerGluGlnAsnLysSerSerProAspProThrSerThrLeuSerAsnThrValSer 2511  
 QY 814 -----GAAATCTCGATAATACA 831  
 Db 2512 TyrGluArgSerThrAspGlySerPheGlnAspArgPheGluGluPheGluAspSerThr 2531  
 QY 832 CTCACGACACCAAGATGCTACACAGAA----- 861  
 Db 2532 LeuLysProAsnArgLysLysProThrGluAsnIleIleAspLeuAspLysGluAsp 2551  
 QY 862 AGAGAAACAGAAATTCGCTGTGGAGGACGAGGGTGGAGCTCAGCGTCTCTCTGTA 921  
 Db 2552 LysAspLeuIleLeuThrIleThrGluSerThrIleLeuGluIleLeuProGluLeuThr 2571  
 QY 922 AACCAAGTTCAAGCGAGCTGCTGCTCCAGTCCCATATTTACCAGGAGCTAGCA 981  
 Db 2572 SerAspLysAsnThrIleIleAspIleAspHisThrLysProValTyrGluAspIleLeu 2591  
 QY 982 GGAAAGTCCCACTTCAGATGCAAAAGATATTTAGAAACTTCAGAGTTCAAAAATC 1041  
 Db 2592 Gly-----MetGlnThrAspIleAspThrGluValProSer----- 2603  
 QY 1042 CATGTGTTAGGATTTAGACCAAGAAAGAAAGAGTGGCTCAAGCTCCACAGATGCAA 1101  
 Db 2604 -----GluProHisAspSerAsnAspGluSerAsnAspSerThrGln 2618  
 QY 1102 CTTAGGGCCATCTTTAAGAGACACAGTGCAGAGCAAAAAGCCCTCCAAAGTGAACCTCTG 1161  
 Db 2619 ValGlnGluIleTyrGlu-----AlaAlaValAsnLeu 2629  
 QY 1162 TCTTTTGAATCCAAACAAATTTGAAGTGAAGTCAATCATCGAACCAATGGAGGAGAC 1221  
 Db 2630 SerLeu-----ThrGluGluThrPheGluGlySer----- 2639  
 QY 1222 AAGCAACCAAGAAATCTATCTCAAGCTACAGACTCAAAAGGCTGATCAGCAAGCACTA 1281

Db 2640 -----AlaaspValLeuAlaSerTyrThrGlnAlaThrHisAspGluSerMetThrTyr 2657  
 QY 1282 GAGGAAGACAATCTTTGGATGTGGGGACAATTCAGATTCACTGATGAAATTCGTGGATCA 1341  
 Db 2658 GluAspArgSerGlnLeuAspHisMetGlyPheHisPheThrThr-----Gly 2673  
 QY 1342 CTGCCAGCTTTGGTCTCGACACCAATCAGAGCTG-----CCACA 1383  
 Db 2674 IleProAla-----ProSerThrGluThrGluLeuAspValLeuLeuProThrAlaThr 2691  
 QY 1384 TCTTTTGTCTGTTATAACAGAGGATGCTACTTTGAGTCACAGACTTCTCCCTGTGACCC 1443  
 Db 2692 SerLeuProIleProArgLysSerAlaThrValIleProGluIleGluIleLysAla 2711  
 QY 1444 CAGCTTGACAGACGTGAC----- 1461  
 Db 2712 GluAlaLysAlaLeuAspMetPheGluSerSerThrLeuSerAspGlyGlnAlaIle 2731  
 QY 1461 ----- 1461  
 Db 2732 AlaAspGlnSerGluIleIleProThrLeuGlyGlnPheGluArgThrGlnGluGluTyr 2751  
 QY 1462 -----GGAGCAGAGCAT 1473  
 Db 2752 GluAspLysLysHisAlaGlyProSerPheGlnProGluPheSerSerGlyAlaGluGlu 2771  
 QY 1474 GGTCTA-----CCTGACACTTCTGTGCTCCACCTGCTATGGCTCTACCTCC 1521  
 Db 2772 AlaLeuValAspHisThrProTyrLeuSerIleAlaThrThrHisLeuMetAspGlnSer 2791  
 QY 1522 CTGTCAAGAGCT-----CCACCTTCTTT----- 1545  
 Db 2792 ValThrGluValProAspValMetGluGlySerAsnProProTyrTyrThrThrThr 2811  
 QY 1546 ATGGCATCAAGCATCTTCTCTCTGATCAAGGACCAACAGATACATGGCCACT----- 1602  
 Db 2812 LeuAlaValSerThrPheAlaLysLeuSerSerGluThrProSerSerProLeuThrIle 2831  
 QY 1603 -----GACCAGACATGCTAGTACGAGG 1626  
 Db 2832 TyrSerGlySerGluAlaSerGlyHisThrIleProGlnProSerAlaLeuProGly 2851  
 QY 1627 CTCACCATCCCCCAGT----- 1644  
 Db 2852 IleAspValGlySerSerValMetSerProGlnAspSerPheLysGluIleHisValAsn 2871  
 QY 1645 -----GATTATTCTGCAATCGCAACTGGCTCTGGGAATTTCAATCCACTGCA 1695  
 Db 2872 IleGluAlaThrPheLysProSerSerGluGluTyrLeuHisIleThrGluProSer 2891  
 QY 1696 TCTTCAGATGACAGCCGATCAAGTCCAGGTGGCGAGAT-----ATGTCACAGACCTAGAT 1752  
 Db 2892 LeuSerProAspThrLysLeuGluProSerGluAspGlyLysProGluLeuLeuGlu 2911  
 QY 1753 GAAATGGATCTGTCTGACACTCTCTGCC----- 1779  
 Db 2912 GluMetGluAlaSerProThrGluLeuIleAlaValGluGlyThrGluIleLeuGlnAsp 2931  
 QY 1780 -----CCATCT----- 1785  
 Db 2932 PheGlnAsnLysThrAspGlyGlnValSerGlyGluAlaIleLysMetPheProThrIle 2951  
 QY 1786 GAGGTACCAAGCTCAGCGAATATGTTCTGTGCCAGATCATTC-----TTGAGAGATACC 1842  
 Db 2952 LysThrProGluAlaGlyThrValIleThrThrAlaAspGluIleGluLeuGlyAla 2971  
 QY 1843 ACT----- 1845  
 Db 2972 ThrGlnTrpProHisSerThrSerAlaSerAlaThrTyrGlyValGluAlaGlyValVal 2991  
 QY 1846 -----CCTGTCTCAGCTTTACAGTATATACCATAGTTCTTATGACCATGGCC 1893



Db 2992 ProTyrLeuSerProGlnThrSerGluArgProThrLeuSerSerSerProGluLeuLeu 3011  
QY 1894 CCC-----AAGGCCGAGAGCTGTAGTCTTCTTCTCAGTCTGCT 1932  
Db 3012 ProGluThrGlnAlaAlaLeuLeuLeuLeuLeuLeuLeuLeuLeu 3031  
QY 1933 -----GTTGCTACATGCGCTTCTCCACAGCAGCTGTCAACAGAGCTCTCTGGAGTAC 1986  
Db 3032 GlnValAlaAlaAlaGluLeuLeuLeuLeuLeuLeuLeuLeuLeu 3051  
QY 1987 CGAGCTTGGAGCAAAATTCACAGCAGCTGCTGCTTCCATATCTA-----CGA 2034  
Db 3052 Asn-----ThrGluValAlaThrProPheSerLeuLeuLeuLeu 3065  
QY 2035 TCCAACTTTACAGGATTTAAGCAACTTGAATATCTTCACTTCAAGAACGGAGTGTG--- 2091  
Db 3066 SerAsnGluThrAspPhe-----LeuIleGlyLeuLeuLeuLeuLeu 3081  
QY 2092 -----ATTGTGAATAGC 2103  
Db 3082 GlyThrAlaIleTyrLeuProGlyProAspArgCysLysMetAsnProCysLeuLeuLeu 3101  
QY 2104 AAATGAAGTTTGTAGTCTGTCGCTGATACCTCACCAGGCTGTGCACGGCTCTTG 2163  
Db 3102 GlyThrCysTyrProThrGluThrSerTyrValCysThr---CysValProGlyTyrSer 3120  
QY 2164 GAGGATTTTGTCTGCTGTCAGCCCAACACTCCATCTGGAATATAGACAGCTACTCTCTC 2223  
Db 3121 GlyAsp-----GlnCysGluLeuLeuLeuLeuLeuLeu 3130  
QY 2224 AACATTGACAGCTGATCAAGCAGATCCCTGCAAGTTCTTGGCTGCGGCGAATTTGCC 2283  
Db 3131 -----CysHisSerAsnProCysArgAsnGlyAla 3140  
QY 2284 CAATGCTTAAGAACGAACGAGCTGAGGAGCGAGTCTGCTGCAACAGGATATGAC 2343  
Db 3141 ThrCysValAspGlyPheAsnThr-----PheArgCysLeuLeuLeuLeuLeuLeu 3158  
QY 2344 AGCCAGGAGGAGCTGAGCGTCTGGAACAGCGCTTCTGCTGCGCTGCAACAGGATATGAC 2403  
Db 3159 Gly-----AlaLeuCysGluLeuLeuLeuLeuLeuLeu 3169  
QY 2404 GAGTCT-----CTCCAGGGAAG-----CGAGCTCCATGAGGCTG 2439  
Db 3170 AspTyrGlyTyrPheHisLeuGlnGlyCysTyrTyrTyrPheAlaHisArgArgThr 3189  
QY 2440 CCAGATCACTCGAATCAAA-----GCATACAAACTAGTGTAAAG 2484  
Db 3190 TrpAspAlaAlaGluArgGluCysArgLeuGlnGlyAlaHisLeuThrSerIleLeuSer 3209  
QY 2485 TTCAAAATCAACAA-----AAT 2502  
Db 3210 HisGluGluGlnMetPheValAsnArgValGlyHisAspTyrGlnTrpIleGlyLeuLeu 3229  
QY 2503 AACAGGTAATCAGTAA-----AGAAATTCGAATTAAGTACCTGAGTATGAGTAA 2556  
Db 3230 AspLysMetPheGluHisAspPheArgTyrThrAspGlySerThrLeuGlnTyrGluLeu 3249  
QY 2557 TTT-----ACCATCAAGT----- 2571  
Db 3250 TrpArgProAsnGlnProAspSerPheSerAlaGlyGluAspCysValIleLeu 3269  
QY 2572 TGGGAAGGAAATTAATAACAGTAA-----ATGTACAAATTAATCACTTAGCTAT---CTC 2622  
Db 3270 TrpHisGluLeuGlnTrpAsnAspValProCysAsnTyrHisLeuTyrThrCys 3289  
QY 2623 AAGAGAGATGATTG-----CCTTCTCAAGGAAATGACAGAGGATAT 2667  
Db 3290 LysLeuGlyThrValAlaCysGlyGlnProProValValGluLeuLeu-----LysThr 3307  
QY 2668 TCATGGGTCATCAAAATCAGACATACAGTACAGTACAGTACAGTACAGTACATATTT 2727  
Db 3307 rPheGlyLysMetLysProArgTyrGluLeuLeuSer----- 3319

QY 2728 CAATATAGAGAGTCATGCTACTTGGCAACAGTAATTTCTGAAAAAAGACACTTACT 2787  
Db 3320 -----LeuIleArgTyrHisCysLysAspGlyPhe---IleGlnArgHisLeuPr 3335  
QY 2788 TATTATTAAA-----ACCC 2802  
Db 3335 oThrIleArgCysLeuGlyAsnGlyArgTrpAlaIleProLysIleThrCysMetAsnPr 3355  
QY 2803 AAATGCAATCAGCGAAACATAT 2824  
Db 3355 oSerAlaTyrGlnArgThrTyr 3362  
RESULT 13  
NIBA HUMAN  
ID NIBA HUMAN STANDARD; PRT; 928 AA.  
AC Q9BQ8; Q8TIS; Q9H9Y8; Q9H9Y8; Q9HCB9;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Niban protein.  
GN CLOPF24.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE=20546198; PubMed=11011112;  
RA Majima S., Kajino K., Fukuda T., Otsuka F., Hino O.;  
RT "A novel gene 'Niban' upregulated in renal carcinogenesis: cloning by  
RT the cDNA-amplified fragment length polymorphism approach.";  
RL Jpn. J. Cancer Res. 91:869-874 (2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21218927; PubMed=11318611;  
RA Sood R., Bonner T.I., Malakowaka I., Stephan D.A., Robbins C.M.,  
RA Connors T.D., Morgenbesser S.D., Su K., Faruque M.U., Pinkett H.,  
RA Graham C., Baxevas A.D., Klinger K.W., Landes G.M., Trent J.M.,  
RA Carpen J.D.;  
RT "Cloning and characterization of 13 novel transcripts and the human  
RT RS8 gene from the 1q25 region encompassing the hereditary prostate  
RT cancer (hpc1) locus.";  
RL Genomics 73:211-222 (2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,  
RA Ninomiya K., Iwayanagi T.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Blood;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner K.H., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Heiton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

|    |      |   |                          |                  |
|----|------|---|--------------------------|------------------|
| QY | 551  | ----  | -----GGAAAA              | 556              |
| Db | 245  | ValMetGluGluLeuLeuProThrLeuGlnThrAspLeuLeuProLysMetLysGlyLys  | 264                      |                  |
| QY | 557  | AACTTCAGCAATTC  | CCAGGACACCTGCATCTTCTCCAG | 598              |
| Db | 265  | LysAsnAspArgLysArgThrTrpLeuGlyLeuGluGluAlaTyrThrLeuValGln     | 284                      |                  |
| QY | 599  | AGAAATAAACAGAGAAGTTTCCCTCACAGAAAAGATGAATATCTGCA               | 646                      |                  |
| Db | 285  | HisGlnValSerGluGlyLeuSerAlaLeuLysGluGluCysArgAlaLeuThrLysGly  | 304                      |                  |
| QY | 647  | ---CAGAACACATTGGGAGAGCTGGTGAACACCATGTCTATTCAACAGCAATCTCAATT   | 703                      |                  |
| Db | 305  | LeuGluGlyThrIleArgSerAspMetAspGlnIleValAsnSerLysAsnTyrLeuIle  | 324                      |                  |
| QY | 704  | TCAAGAGCTGGGAGTATCTTAAGAAAACCCCTCAGAA                         | 751                      | -----GAGCAAAATTC |
| Db | 325  | GlyLys--IleLysAlaMetValAlaGlnProAlaGluLysSerCysLeuGluSerValG  | 344                      |                  |
| QY | 752  | AA-----GATGTTGCCAACGCTCTCACTTGGGCCT                           | 780                      | -----            |
| Db | 344  | lnProPheLeuAlaSerIleLeuGluGluLeuMetGlyProValSerSerGlyPheSerG  | 364                      |                  |
| QY | 781  | -----   | TTCCCTCTCTACTC           | 793              |
| Db | 364  | luValArgValLeuPheGluLysGluValAsnGluValSerGlnAenPheGlnThrThrL  | 384                      |                  |
| QY | 794  | CTGATGACACCCCTCTCAATGAATTTCTCGATAATACACTCAAC                  | 841                      | -----GACA        |
| Db | 384  | ysAspSerValGlnLeuLysGluHisLeuAspArgLeuMetAsnLeuProLeuHisSerV  | 404                      |                  |
| QY | 842  | CCAGAGTCCCTACAAAGAAAGAACAGATTCCTGTGTGGAGAGCAGAGGGTGG          | 901                      | -----            |
| Db | 404  | allysMetGluProCys-----TyrThrLysValAsnLeuLeuHisGluLeuGlnA      | 422                      |                  |
| QY | 902  | AGCTCAGC-----GTCTCTCTGCTA                                     | 931                      | -----AACCAAGACT  |
| Db | 422  | spLeuLysSerArgPheArgPheProHisLeuAspLeuValGlnArgThrGlnAsnT     | 442                      |                  |
| QY | 932  | TCAAGGCGAGCTCGCTGACTCCGAG                                     | 967                      | -----TCGCCATATT  |
| Db | 442  | YrMetGlnGluLeuMetGluAsnAlaValPheThrPheGluGlnLeuLeuSerProHisL  | 462                      |                  |
| QY | 968  | ACCAGCAGCTAGCAGGAAGTCCCACTTCAGTGCAAAAGATATTTAAGAACTTCGAG      | 1027                     | -----            |
| Db | 462  | eugGlnGlyAlaSerLysThrAlaValAlaIleGluLysVal                    | 476                      | -----            |
| QY | 1028 | GATTCAAAAATATCCATGTGTAGGATTTAGACCAAGAAAGAAAGATGGCTCAAGCT      | 1087                     | -----            |
| Db | 477  | -----LysLeuArgValLeu-----LysGlnTyrAspTyrAspSerS               | 499                      |                  |
| QY | 1088 | CCACA-----GAGATGCAATCTACGGCCATCT                              | 1114                     | -----            |
| Db | 489  | erThrIleArgLysLysIlePheGlnGluAlaLeuValGlnIleThrLeuProThrValG  | 509                      |                  |
| QY | 1115 | TTAAGACACACGTGACAGCAAAAGCCCTGCAAGTCACTCTCTCTCTTT              | 1167                     | -----            |
| Db | 509  | lnLysAlaLeuAlaSerThrCysLysProGluLeuGlnLysTyrGluGlnPheIlePheA  | 529                      |                  |
| QY | 1168 | -----GATTCACCAAAATGGAAGTGAGGAAGTCTCATGGAACCATGGAGGAGACA       | 1222                     | -----            |
| Db | 529  | laAspHisThrAsnMetIleHisValGluAenValTyr-----GluGluI            | 544                      |                  |
| QY | 1223 | AGCAACACAGAAATCTATCTCAAGCTACAGACTCAAAAGGCTGATCAGCAAGCTAG      | 1282                     | -----            |
| Db | 544  | leLeuHisGlnIleLeuLeuAspGluThr--LeuLysValIleLysGluAlaIleL      | 563                      |                  |
| QY | 1283 | AGGAAGAACCAATCTTTGGATGTGGGACAATTTCACTGATGAATTCGTGATCATC       | 1342                     | -----            |
| Db | 563  | eulLysLysHisAsnLeuPheGluAspAsnMetAlaLeuProSerGlySerValSerSerL | 583                      |                  |
| QY | 1343 | TGCACGCTTTGGTCTCGACAC   | 1365                     | -----            |

Db 583 euThrAspLeuLysProThrGlySerAsnGlnAlaSerProAlaArgAlaSerA 603  
 QY 1365  
 Db 603 laileuProGlyValLeuGlySerGluThrLeuSerAsnGluValPheGlnGluSerG 623  
 QY 1366  
 Db 623 luGluGluLysGlnProGluValProSerLeuAlaLysGlySerLeuSerLeuP 643  
 QY 1396  
 Db 643 roGlyProSerProProAspGlyThrGluGlnValIleSerArgValAspP 663  
 QY 1433 CTGTGTAACCCACCTTGACAGCTGGACGAGGAGGAGCATGGTCTACCTGACACTTCTT 1492  
 Db 663 roValValAsnProValAlaThrGluAspThrAla-----GlyLeuProGlyThrCys 681  
 QY 1493 GGTCTCCACCTGCTATGGCTCTACTCTCTGCTGTCAGAGCTCCA----- 1536  
 Db 681 er-SerGluLeuGluPheGlyGlyThrLeuGluAspGluGluProAlaGlnGluLuprog 701  
 QY 1537 --CCTTTCTTTATGGCATCAAGCATCTCTCTGCTGATCATCAAGGCACACAGATACAA 1594  
 Db 701 luProIleThrAlaSerGlySerLeuLysAlaLeu----- 712  
 QY 1595 TGGCCACTGACACACAATGCTAGTACCAGGGTCCACATCCCATCCACAGGATATTCTG 1654  
 Db 713 -----ArgLysLeuLeuThrAlaSerValGluValProValAspSerAlaProV 729  
 QY 1655 CAATCAGCCAACTGGCTCTGGGAATTTACATCCCATCCGATCTTCAGATGACAGCCGAT 1714  
 Db 729 alMetGluGluAspThrAsnGlySerHisValProGlnGluAsnGluGluGluGlu 749  
 QY 1715 CRAGTGAGTGGCGAGATATGTCAGACACTAGT-----GAATGGATCTGT 1765  
 Db 749 luLysGluProSerGlnAlaAlaAlaIleHisProAspAsnCysGluGluSerGluValS 769  
 QY 1766 CTGACACTCTGCCCATCTCAGGTACCCAG 1797  
 Db 769 erGluArgGluAlaGlnProProCysProGlu 779  
 RESULT 14  
 POPC\_RALSO  
 ID POPC\_RALSO STANDARD; PRT: 1024 AA.  
 AC Q9RS52;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE PopC protein.  
 GN POPC OR RSP0875 OR RS01646.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OG Plasmid megaplasmid.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GMI1000;  
 RX MEDLINE=20253307; PubMed=10792715;  
 RA Quenon M., Timmers A.C.J., Boucher C., Arlat M.;  
 RT "Two novel proteins, PopB, which has functional nuclear localization  
 RT signals, and PopC, which has a large leucine-rich repeat domain, are  
 RT secreted through the Hrp-secretion apparatus of Ralstonia  
 RL solanacearum";  
 RL Mol. Microbiol. 36:261-277(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GMI1000;  
 RX MEDLINE=1123852; PubMed=1123852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Arlat M., Billault A., Brotter P., Camus J.C., Cattolico L.,

Chandler M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,  
 Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
 Signier P., Thebault P., Whalen M., Wincker P., Levy M.,  
 Weissenbach J., Boucher C.A.;  
 "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 Nature 415:497-502(2002).  
 -!- FUNCTION: Probably involved in host-pathogen interactions. May  
 interact with plant target proteins; may modulate a plant signal  
 transduction pathway. Secreted.  
 -!- SUBCELLULAR LOCATION: Secreted.  
 -!- SIMILARITY: Contains 22 leucine-rich (LRR) repeats.  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 EMBL: AJ245811; CAB57879.1; -  
 DR EMBL; AL646081; CAD18026.1; -  
 DR InterPro; IPR001611; LRR.  
 DR Pfam; PF00560; LRR; 11.  
 KW Repeat; Leucine-rich repeat; Plasmid; Complete proteome.  
 FT REPEAT 192 218 LRR 1.  
 FT REPEAT 223 246 LRR 2.  
 FT REPEAT 248 270 LRR 3.  
 FT REPEAT 271 293 LRR 4.  
 FT REPEAT 295 315 LRR 5.  
 FT REPEAT 316 338 LRR 6.  
 FT REPEAT 340 359 LRR 7.  
 FT REPEAT 360 381 LRR 8.  
 FT REPEAT 382 403 LRR 9.  
 FT REPEAT 404 427 LRR 10.  
 FT REPEAT 429 450 LRR 11.  
 FT REPEAT 451 475 LRR 12.  
 FT REPEAT 477 495 LRR 13.  
 FT REPEAT 496 519 LRR 14.  
 FT REPEAT 521 543 LRR 15.  
 FT REPEAT 544 566 LRR 16.  
 FT REPEAT 568 589 LRR 17.  
 FT REPEAT 590 612 LRR 18.  
 FT REPEAT 614 636 LRR 19.  
 FT REPEAT 637 661 LRR 20.  
 FT REPEAT 662 687 LRR 21.  
 FT REPEAT 824 847 LRR 22.  
 FT CONFLICT 198 198 V -> D (IN REF. 1).  
 SQ SEQUENCE 1024 AA; 110929 MW; 27AEC679E9A4EB7D CRC64;  
 Alignment Scores:  
 Pred. No.: 0.169 Length: 1024  
 Score: 145.50 Matches: 166  
 Percent Similarity: 33.87% Conservative: 106  
 Best Local Similarity: 20.67% Mismatches: 257  
 Query Match: 2.48% Indels: 275  
 DB: 1 Gaps: 33  
 US-10-007-270-1 (1-3330) x POPC\_RALSO (1-1024)  
 QY 675 CCATGTGTCATTTCAACAGCAATCTCATTTCAAGACTTGGCGAGTATTCTAAGAAACC 734  
 Db 6 ProGlnThrValProGlyHisSerProPheTrpProLeuPhePhePheThrHisAsnAsn 25  
 QY 735 CTCAGAGAGCGAATTCAAGATCTGCCACGTCTCACTTGGGCGCTTTCCCTCTCTACTCC 794  
 Db 26 LysArgGluAlaArg-HisAspLeuAlaArgLeuSerLeuThrLeuMetProIleLeuPr 45  
 QY 795 TGATGACACCCCTCTCAATCAATTCGATAATACACTCAACGACACACAGATGCTTAC 854  
 Db 45 OAAG-----LeuPheHisArgThrSerArgThrSerSerAlaAs 58  
 QY 855 AACAGAAAGAGAAACA-----GAATTCGCTGT 881

Db 58 pThrGlnArgAspAlaThrProProAsnAlaSerProLeuHisGlyGluProGlyAr 78  
 QY 882 GTTGGAGGACAGAGGGTGGAGCTCAGCGTCTCTCTGTAAACACAGAGTTCAGGCAGA 941  
 Db 78 gThrProArgSerArgGlyGluGlnArgAsnLeuArg-LeuArgSerAsnAlaGlnr 98  
 QY 942 GCTCGCTGACCTCCAGT----- 958  
 Db 98 hrSerGlyThrProGlyThrProAlaArgProGlnIleArgAlaSerAlaSerArgThrA 118  
 QY 959 -----CCCCATATTACAGAGCTAGCAGCAAGAGTCCCACTTCA 998  
 Db 118 laProSerThrProGlnHisProGlnGlyThrGlu-GlyThrArgThrValProAsnSer 137  
 QY 999 -----GATGCAAAAGATATTAAAGAACTTCCAGGATCAAAAATAATCCATGT 1046  
 Db 138 ProLeuHisAsnAspAla-ArgVal----- 145  
 QY 1047 GTTAGGATTTAGACCAAGAAAGAAAGATGGCTCAAGCTCCACAGAGTGCACACTTAC 1106  
 Db 146 -----PheArgGluArgAlaAspHisThrGlyLeuSerAlaTrpArgThrGluMetLe 163  
 QY 1107 GGCATCTTTAAGAGACACAGTGCAGAGCAAAAGCCCTGCAAGTGACCTCTGTCTTT 1166  
 Db 163 uThrArgPheIleGluHisSerArglys----- 172  
 QY 1167 TGATTCACAAATTTGAAAGTGAGGAGTCTATCATGGAACCAATGGAGGAGGACAAGCA 1226  
 Db 173 -----HisGlyLeuAlaAsnAppPheGluGl 181  
 QY 1227 ACCGAAATCTATCTCACAGCTACAGACCTCAAAAGGTGTATCAGCAAGCACTAGAGGA 1286  
 Db 181 nValArgValTr-----AspArgLeuSerArgAlaVal----- 192  
 QY 1287 AGAACATCTTTGGATGGGGGACAAATTCAGTTCACTGATGAATTCGTGGATCACTGCC 1346  
 Db 193 -----AspHisLeuLysSerValLeuAr 200  
 QY 1347 AGCCTTTGGTCTGACACCCAA---TCAGAGTGGCCACATCTTTTGGTGTATTAAACAGA 1403  
 Db 200 gMetSerGlyAspSerValGlnLeuLysSerLeuProVal----- 213  
 QY 1404 GGATGCTACTTTGAGTCAGAACTTCTCTCTGTGAA-----CCCCA 1445  
 Db 214 -----ProGluLeuProAspValThrPheGluIleAlaHisLeuLysAs 228  
 QY 1446 GCTTGAGACAGTGGAC----- 1461  
 Db 228 nLeuGluThrValAspCysAspLeuHisAlaLeuProAlaThrLeuGluAsnLeuPheLe 248  
 QY 1462 -----GGAGCAGAGCAT-----GGTCTACTGTGACACTTCTTG 1493  
 Db 248 uLeuGluThrLeuSerLeuLysGlyAlaLysAsnPheLysAlaLeuProAspAlaValTr 268  
 QY 1494 GTCTCCACTGTATGGCTCTACTCTCTGTGCAAGCT-----CCACC 1538  
 Db 268 pArgLeuProAlaLeuGlnLeuLysLeuSerGluThrGlyLeuLysSerLeuProPr 288  
 QY 1539 TTTCTTTATGGCATCAGCATCTTCTCTGACTGATCAAGGCACACAGATACATGGC 1598  
 Db 288 oValGlyGlySerAlaLeuGlnArgLeuThrIleGluAspSer----- 303  
 QY 1599 CACTGACACAGACATGCTAGTACCGGGCTCACCTCCACC-----ATGATTTATTC 1652  
 Db 304 -----ProLeuGluGlnLeuProAlaGlyPheAlaAspLeuAs 316  
 QY 1653 TGCAATCAGCAACTGGCTCTGGGAATTTTCATCCACCTGCACTTTCAGATGACAGCCG 1712  
 Db 316 pGlnLeuAlaSerLeuSerLeu-----SerAsnThrLysLeuGl 329  
 QY 1713 ATCAAGTGCAGGTGGCGAAGATATGGTCAGACACCTAGATGAATGGATCTGTGACAC 1772

Db 329 uLysLeuSerSerGlyIleGlyGlnLeuProAlaLeuLysSerLeuSerLeuGlnAspAs 349  
 QY 1773 TCCT-----GCCCATCTCAGGTACCGAGACTCAGCAATATGTTTCTGTGCC 1820  
 Db 349 nProLysLeuGluArgLeuProLysSerLeuGlyGlnValGluGluLeuThrLeuIleGl 369  
 QY 1821 AGAT-----CATTTCTTGCGATACACCTCTCTCAGCTTTTACAGTATATCACCAC 1874  
 Db 369 yGlyArgIleHisAlaLeuProSerAlaSerGlyMetSerSerLeuGlnLysLeuThrVa 389  
 QY 1875 TAGTTCTATGACCAATGCCCCCAAGGCGAGAGCTGGTAGTGTCTTCTTCACTGTGGTGT 1934  
 Db 389 lAspAsnSerSerLeuAlaLysLeuProAlaAspPheGlyAlaLeuGlyAsnLeu----- 407  
 QY 1935 TGCTAACATGGCTTCTCCAAC-----GACCTGTTCAACAA 1970  
 Db 408 -AlaHisValSerLeuSerAsnThrLysLeuArgAspLeuProAlaSerIleGlyAsnLe 427  
 QY 1971 GAGCTCTCTGAGTACCGAGCTCTGAGCAACAATTCACAGCTGCTGGTTCATATCT 2030  
 Db 427 uPheThrLeuLysThrLeuSerLeuGlnAspAsnProLysLeuGlySerLeuProAlaSe 447  
 QY 2031 ACATCCATCTTACAGATTTTAAGCAACTTGAATCTTAACTTCAGAAACGGAGTGT 2090  
 Db 447 rPheGlyGlnLeuSerGlyLeuGlnGluLeuThrLeu-----AsnGlyAsnAr 463  
 QY 2091 GATTGTGAATAGCAAAATGAATGTTGCTAAGTCTGTGCGGTATAACCTCAC----- 2142  
 Db 463 gIleHisGluLeuProSerMetGlyGlyAlaSerSerLeuGlnThrLeuThrValAspAs 483  
 QY 2143 -AAGGCTGTGACGGGCTCTGAGGATTTTCTCTGTCGAGCCCAACAACCTCATCT 2201  
 Db 483 pThrAlaLeuAlaGlyLeuProAlaAspPheGlyAla-LeuArgAsn----- 498  
 QY 2202 GGAATACAGACGACTACTCTCTCACTTGAACCCAGCTGATCAGCAGATCCCTGCAAGT 2261  
 Db 499 -----LeuAlaHisLeuSerSerAsnThrGlnLeuArgGluLeuProAlaAsn 516  
 QY 2262 CTGCGCTGCGGCAATTTGCCAATGTGTAAAGAACGACGAGTCTGAGGACGGAGTGT 2321  
 Db 516 hr-----GlyAsnLeu-----HisAlaLeuLysThrLeuSerL 527  
 QY 2322 TCGCTGCAACACAGATATGAC-----AGCCAGGCGAGCTGCGAGCTCTGG 2368  
 Db 527 euGlnGlyAsnGlnLeuAlaThrLeuProSerSerLeuGlyTyrlSerGlyLeuG 547  
 QY 2369 AA-----CCAGGCTCTGTGGCCCTGGCA 2392  
 Db 547 lGluLeuThrLeuLysAsnSerSerValSerGluLeuProProMet---GlyProGlyS 566  
 QY 2393 CAAGAATCGAGTCTCTCCAGGGAAGGGAGCTCCA----- 2430  
 Db 566 er---AlaLeuLysThrLeuThrValGluAsnSerProLeuThrSerIleProAlaAspI 585  
 QY 2431 -----TCAGGTGGCAGATCCTCTGTAAATCAAGCATACAAAATAGTGTGA 2479  
 Db 585 leGlyIleGlnCysGluArg-LeuThrGlnLeuSerLeuSerAsnThrGlnLeuArgAla 604  
 QY 2480 AAAAGTTCCAAATCAACAAATAACAGGTAACTCAGTAAAGAAATCTGATTTACTGA 2539  
 Db 605 LeuProSerSerIleGlyLysLeuSer----- 613  
 QY 2540 CCGTAGAATATCAAGAAATTTAAACATCAAGATTGGAGGAATTTAAACATGAAATGT 2599  
 Db 614 -----AsnLeuLysGlyLeuThrLeuLysAsnAsnAlaArgLeuGlu----- 627  
 QY 2600 ACAATTATCATAGGTATCTCAGAGAGATGATTTGCCTTCTCAGGAAATGGAGAC 2659  
 Db 628 -----LeuLeuSerGluSerGlyVal 634  
 QY 2660 AGGCATATTCTGGGTATCAAAATCCAGACATACAGTCAACCTGAGGAATCAGC 2714  
 Db 635 ArgLysLeuGluSerValArgLysIle-AspLeuSerGlyCysValArgLeuThr 652

RESULT 15  
MUC1\_MOUSE  
ID MUC1\_MOUSE STANDARD; PRT: 630 AA.  
AC Q02496;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Mucin 1 precursor (Polymorphic epithelial mucin) (PEMT) (Episialin).  
DN MUC1 OR MUC-1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
[1] \_SEQUENCE FROM N.A.  
RN MEDLINE=91332029; PubMed=1714452;  
RX Spicer A.P., Parry G., Patton S., Gendler S.J.;  
RT "Molecular cloning and analysis of the mouse homologue of the tumor-associated mucin, MUC1, reveals conservation of potential O-glycosylation sites, transmembrane, and cytoplasmic domains and a loss of minisatellite-like polymorphism.";  
RT J. Biol. Chem. 266:15099-15109(1991).  
RN [2]  
RN \_SEQUENCE FROM N.A.  
RX MEDLINE=92068178; PubMed=1958179;  
RA Vos H.L., Devries V., Hilkens J.;  
RT "The mouse episialin (Muc1) gene and its promoter: rapid evolution of the repetitive domain in the protein.";  
RT Biochem. Biophys. Res. Commun. 181:121-130(1991).  
RL CC -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN  
CYTOSKELETON.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. EXCLUSIVELY LOCATED IN THE APICAL DOMAIN OF THE PLASMA MEMBRANE OF HIGHLY POLARIZED EPITHELIAL CELLS.  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF EPITHELIAL TISSUES. ABERRANTLY EXPRESSED IN EPITHELIAL CARCINOMAS.  
CC -!- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.  
CC -!- SIMILARITY: Contains 1 SEA domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M84683; AAA39756.1; -;  
CC EMBL; U16175; AAA98538.1; -;  
CC EMBL; M65132; AAA39755.1; -;  
CC EMBL; M64928; AAA39755.1; JOINED.  
CC EMBL; M77226; AAA39754.1; -;  
CC PIR; A39344; A39344.  
CC FIC; 152257; 152257.  
CC MGD; MGI:97231; Muc1.  
CC InterPro; IPR000082; SEA\_domain.  
CC Pfam; PF01390; SEA; 1.  
CC SMART; SM00200; SEA; 1.  
CC PROSITE; PS50024; SEA; 1.  
KW Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane; Repeat.  
FT SIGNAL 1 20  
FT CHAIN 21 630  
FT DOMAIN 21 535  
FT TRANSMEM 536 556  
FT DOMAIN 557 630  
FT DOMAIN 42 366  
FT REPEAT 42 61  
FT REPEAT 62 81  
FT REPEAT 82 101  
FT REPEAT 102 122  
FT REPEAT 123 143

144 REPEAT 144 164  
145 REPEAT 145 184  
146 REPEAT 146 204  
147 REPEAT 147 225  
148 REPEAT 148 246  
149 REPEAT 149 266  
150 REPEAT 150 286  
151 REPEAT 151 306  
152 REPEAT 152 326  
153 REPEAT 153 346  
154 REPEAT 154 366  
155 REPEAT 155 386  
156 REPEAT 156 406  
157 REPEAT 157 426  
158 REPEAT 158 446  
159 REPEAT 159 466  
160 REPEAT 160 486  
161 REPEAT 161 506  
162 REPEAT 162 526  
163 REPEAT 163 546  
164 REPEAT 164 566  
165 REPEAT 165 586  
166 REPEAT 166 606  
167 REPEAT 167 626  
168 REPEAT 168 646  
169 REPEAT 169 666  
170 REPEAT 170 686  
171 REPEAT 171 706  
172 REPEAT 172 726  
173 REPEAT 173 746  
174 REPEAT 174 766  
175 REPEAT 175 786  
176 REPEAT 176 806  
177 REPEAT 177 826  
178 REPEAT 178 846  
179 REPEAT 179 866  
180 REPEAT 180 886  
181 REPEAT 181 906  
182 REPEAT 182 926  
183 REPEAT 183 946  
184 REPEAT 184 966  
185 REPEAT 185 986  
186 REPEAT 186 1006  
187 REPEAT 187 1026  
188 REPEAT 188 1046  
189 REPEAT 189 1066  
190 REPEAT 190 1086  
191 REPEAT 191 1106  
192 REPEAT 192 1126  
193 REPEAT 193 1146  
194 REPEAT 194 1166  
195 REPEAT 195 1186  
196 REPEAT 196 1206  
197 REPEAT 197 1226  
198 REPEAT 198 1246  
199 REPEAT 199 1266  
200 REPEAT 200 1286  
201 REPEAT 201 1306  
202 REPEAT 202 1326  
203 REPEAT 203 1346  
204 REPEAT 204 1366  
205 REPEAT 205 1386  
206 REPEAT 206 1406  
207 REPEAT 207 1426  
208 REPEAT 208 1446  
209 REPEAT 209 1466  
210 REPEAT 210 1486  
211 REPEAT 211 1506  
212 REPEAT 212 1526  
213 REPEAT 213 1546  
214 REPEAT 214 1566  
215 REPEAT 215 1586  
216 REPEAT 216 1606  
217 REPEAT 217 1626  
218 REPEAT 218 1646  
219 REPEAT 219 1666  
220 REPEAT 220 1686  
221 REPEAT 221 1706  
222 REPEAT 222 1726  
223 REPEAT 223 1746  
224 REPEAT 224 1766  
225 REPEAT 225 1786  
226 REPEAT 226 1806  
227 REPEAT 227 1826  
228 REPEAT 228 1846  
229 REPEAT 229 1866  
230 REPEAT 230 1886  
231 REPEAT 231 1906  
232 REPEAT 232 1926  
233 REPEAT 233 1946  
234 REPEAT 234 1966  
235 REPEAT 235 1986  
236 REPEAT 236 2006  
237 REPEAT 237 2026  
238 REPEAT 238 2046  
239 REPEAT 239 2066  
240 REPEAT 240 2086  
241 REPEAT 241 2106  
242 REPEAT 242 2126  
243 REPEAT 243 2146  
244 REPEAT 244 2166  
245 REPEAT 245 2186  
246 REPEAT 246 2206  
247 REPEAT 247 2226  
248 REPEAT 248 2246  
249 REPEAT 249 2266  
250 REPEAT 250 2286  
251 REPEAT 251 2306  
252 REPEAT 252 2326  
253 REPEAT 253 2346  
254 REPEAT 254 2366  
255 REPEAT 255 2386  
256 REPEAT 256 2406  
257 REPEAT 257 2426  
258 REPEAT 258 2446  
259 REPEAT 259 2466  
260 REPEAT 260 2486  
261 REPEAT 261 2506  
262 REPEAT 262 2526  
263 REPEAT 263 2546  
264 REPEAT 264 2566  
265 REPEAT 265 2586  
266 REPEAT 266 2606  
267 REPEAT 267 2626  
268 REPEAT 268 2646  
269 REPEAT 269 2666  
270 REPEAT 270 2686  
271 REPEAT 271 2706  
272 REPEAT 272 2726  
273 REPEAT 273 2746  
274 REPEAT 274 2766  
275 REPEAT 275 2786  
276 REPEAT 276 2806  
277 REPEAT 277 2826  
278 REPEAT 278 2846  
279 REPEAT 279 2866  
280 REPEAT 280 2886  
281 REPEAT 281 2906  
282 REPEAT 282 2926  
283 REPEAT 283 2946  
284 REPEAT 284 2966  
285 REPEAT 285 2986  
286 REPEAT 286 3006  
287 REPEAT 287 3026  
288 REPEAT 288 3046  
289 REPEAT 289 3066  
290 REPEAT 290 3086  
291 REPEAT 291 3106  
292 REPEAT 292 3126  
293 REPEAT 293 3146  
294 REPEAT 294 3166  
295 REPEAT 295 3186  
296 REPEAT 296 3206  
297 REPEAT 297 3226  
298 REPEAT 298 3246  
299 REPEAT 299 3266  
300 REPEAT 300 3286  
301 REPEAT 301 3306  
302 REPEAT 302 3326  
303 REPEAT 303 3346  
304 REPEAT 304 3366  
305 REPEAT 305 3386  
306 REPEAT 306 3406  
307 REPEAT 307 3426  
308 REPEAT 308 3446  
309 REPEAT 309 3466  
310 REPEAT 310 3486  
311 REPEAT 311 3506  
312 REPEAT 312 3526  
313 REPEAT 313 3546  
314 REPEAT 314 3566  
315 REPEAT 315 3586  
316 REPEAT 316 3606  
317 REPEAT 317 3626  
318 REPEAT 318 3646  
319 REPEAT 319 3666  
320 REPEAT 320 3686  
321 REPEAT 321 3706  
322 REPEAT 322 3726  
323 REPEAT 323 3746  
324 REPEAT 324 3766  
325 REPEAT 325 3786  
326 REPEAT 326 3806  
327 REPEAT 327 3826  
328 REPEAT 328 3846  
329 REPEAT 329 3866  
330 REPEAT 330 3886  
331 REPEAT 331 3906  
332 REPEAT 332 3926  
333 REPEAT 333 3946  
334 REPEAT 334 3966  
335 REPEAT 335 3986  
336 REPEAT 336 4006  
337 REPEAT 337 4026  
338 REPEAT 338 4046  
339 REPEAT 339 4066  
340 REPEAT 340 4086  
341 REPEAT 341 4106  
342 REPEAT 342 4126  
343 REPEAT 343 4146  
344 REPEAT 344 4166  
345 REPEAT 345 4186  
346 REPEAT 346 4206  
347 REPEAT 347 4226  
348 REPEAT 348 4246  
349 REPEAT 349 4266  
350 REPEAT 350 4286  
351 REPEAT 351 4306  
352 REPEAT 352 4326  
353 REPEAT 353 4346  
354 REPEAT 354 4366  
355 REPEAT 355 4386  
356 REPEAT 356 4406  
357 REPEAT 357 4426  
358 REPEAT 358 4446  
359 REPEAT 359 4466  
360 REPEAT 360 4486  
361 REPEAT 361 4506  
362 REPEAT 362 4526  
363 REPEAT 363 4546  
364 REPEAT 364 4566  
365 REPEAT 365 4586  
366 REPEAT 366 4606  
367 REPEAT 367 4626  
368 REPEAT 368 4646  
369 REPEAT 369 4666  
370 REPEAT 370 4686  
371 REPEAT 371 4706  
372 REPEAT 372 4726  
373 REPEAT 373 4746  
374 REPEAT 374 4766  
375 REPEAT 375 4786  
376 REPEAT 376 4806  
377 REPEAT 377 4826  
378 REPEAT 378 4846  
379 REPEAT 379 4866  
380 REPEAT 380 4886  
381 REPEAT 381 4906  
382 REPEAT 382 4926  
383 REPEAT 383 4946  
384 REPEAT 384 4966  
385 REPEAT 385 4986  
386 REPEAT 386 5006  
387 REPEAT 387 5026  
388 REPEAT 388 5046  
389 REPEAT 389 5066  
390 REPEAT 390 5086  
391 REPEAT 391 5106  
392 REPEAT 392 5126  
393 REPEAT 393 5146  
394 REPEAT 394 5166  
395 REPEAT 395 5186  
396 REPEAT 396 5206  
397 REPEAT 397 5226  
398 REPEAT 398 5246  
399 REPEAT 399 5266  
400 REPEAT 400 5286  
401 REPEAT 401 5306  
402 REPEAT 402 5326  
403 REPEAT 403 5346  
404 REPEAT 404 5366  
405 REPEAT 405 5386  
406 REPEAT 406 5406  
407 REPEAT 407 5426  
408 REPEAT 408 5446  
409 REPEAT 409 5466  
410 REPEAT 410 5486  
411 REPEAT 411 5506  
412 REPEAT 412 5526  
413 REPEAT 413 5546  
414 REPEAT 414 5566  
415 REPEAT 415 5586  
416 REPEAT 416 5606  
417 REPEAT 417 5626  
418 REPEAT 418 5646  
419 REPEAT 419 5666  
420 REPEAT 420 5686  
421 REPEAT 421 5706  
422 REPEAT 422 5726  
423 REPEAT 423 5746  
424 REPEAT 424 5766  
425 REPEAT 425 5786  
426 REPEAT 426 5806  
427 REPEAT 427 5826  
428 REPEAT 428 5846  
429 REPEAT 429 5866  
430 REPEAT 430 5886  
431 REPEAT 431 5906  
432 REPEAT 432 5926  
433 REPEAT 433 5946  
434 REPEAT 434 5966  
435 REPEAT 435 5986  
436 REPEAT 436 6006  
437 REPEAT 437 6026  
438 REPEAT 438 6046  
439 REPEAT 439 6066  
440 REPEAT 440 6086  
441 REPEAT 441 6106  
442 REPEAT 442 6126  
443 REPEAT 443 6146  
444 REPEAT 444 6166  
445 REPEAT 445 6186  
446 REPEAT 446 6206  
447 REPEAT 447 6226  
448 REPEAT 448 6246  
449 REPEAT 449 6266  
450 REPEAT 450 6286  
451 REPEAT 451 6306  
452 REPEAT 452 6326  
453 REPEAT 453 6346  
454 REPEAT 454 6366  
455 REPEAT 455 6386  
456 REPEAT 456 6406  
457 REPEAT 457 6426  
458 REPEAT 458 6446  
459 REPEAT 459 6466  
460 REPEAT 460 6486  
461 REPEAT 461 6506  
462 REPEAT 462 6526  
463 REPEAT 463 6546  
464 REPEAT 464 6566  
465 REPEAT 465 6586  
466 REPEAT 466 6606  
467 REPEAT 467 6626  
468 REPEAT 468 6646  
469 REPEAT 469 6666  
470 REPEAT 470 6686  
471 REPEAT 471 6706  
472 REPEAT 472 6726  
473 REPEAT 473 6746  
474 REPEAT 474 6766  
475 REPEAT 475 6786  
476 REPEAT 476 6806  
477 REPEAT 477 6826  
478 REPEAT 478 6846  
479 REPEAT 479 6866  
480 REPEAT 480 6886  
481 REPEAT 481 6906  
482 REPEAT 482 6926  
483 REPEAT 483 6946  
484 REPEAT 484 6966  
485 REPEAT 485 6986  
486 REPEAT 486 7006  
487 REPEAT 487 7026  
488 REPEAT 488 7046  
489 REPEAT 489 7066  
490 REPEAT 490 7086  
491 REPEAT 491 7106  
492 REPEAT 492 7126  
493 REPEAT 493 7146  
494 REPEAT 494 7166  
495 REPEAT 495 7186  
496 REPEAT 496 7206  
497 REPEAT 497 7226  
498 REPEAT 498 7246  
499 REPEAT 499 7266  
500 REPEAT 500 7286  
501 REPEAT 501 7306  
502 REPEAT 502 7326  
503 REPEAT 503 7346  
504 REPEAT 504 7366  
505 REPEAT 505 7386  
506 REPEAT 506 7406  
507 REPEAT 507 7426  
508 REPEAT 508 7446  
509 REPEAT 509 7466  
510 REPEAT 510 7486  
511 REPEAT 511 7506  
512 REPEAT 512 7526  
513 REPEAT 513 7546  
514 REPEAT 514 7566  
515 REPEAT 515 7586  
516 REPEAT 516 7606  
517 REPEAT 517 7626  
518 REPEAT 518 7646  
519 REPEAT 519 7666  
520 REPEAT 520 7686  
521 REPEAT 521 7706  
522 REPEAT 522 7726  
523 REPEAT 523 7746  
524 REPEAT 524 7766  
525 REPEAT 525 7786  
526 REPEAT 526 7806  
527 REPEAT 527 7826  
528 REPEAT 528 7846  
529 REPEAT 529 7866  
530 REPEAT 530 7886  
531 REPEAT 531 7906  
532 REPEAT 532 7926  
533 REPEAT 533 7946  
534 REPEAT 534 7966  
535 REPEAT 535 7986  
536 REPEAT 536 8006  
537 REPEAT 537 8026  
538 REPEAT 538 8046  
539 REPEAT 539 8066  
540 REPEAT 540 8086  
541 REPEAT 541 8106  
542 REPEAT 542 8126  
543 REPEAT 543 8146  
544 REPEAT 544 8166  
545 REPEAT 545 8186  
546 REPEAT 546 8206  
547 REPEAT 547 8226  
548 REPEAT 548 8246  
549 REPEAT 549 8266  
550 REPEAT 550 8286  
551 REPEAT 551 8306  
552 REPEAT 552 8326  
553 REPEAT 553 8346  
554 REPEAT 554 8366  
555 REPEAT 555 8386  
556 REPEAT 556 8406  
557 REPEAT 557 8426  
558 REPEAT 558 8446  
559 REPEAT 559 8466  
560 REPEAT 560 8486  
561 REPEAT 561 8506  
562 REPEAT 562 8526  
563 REPEAT 563 8546  
564 REPEAT 564 8566  
565 REPEAT 565 8586  
566 REPEAT 566 8606  
567 REPEAT 567 8626  
568 REPEAT 568 8646  
569 REPEAT 569 8666  
570 REPEAT 570 8686  
571 REPEAT 571 8706  
572 REPEAT 572 8726  
573 REPEAT 573 8746  
574 REPEAT 574 8766  
575 REPEAT 575 8786  
576 REPEAT 576 8806  
577 REPEAT 577 8826  
578 REPEAT 578 8846  
579 REPEAT 579 8866  
580 REPEAT 580 8886  
581 REPEAT 581 8906  
582 REPEAT 582 8926  
583 REPEAT 583 8946  
584 REPEAT 584 8966  
585 REPEAT 585 8986  
586 REPEAT 586 9006  
587 REPEAT 587 9026  
588 REPEAT 588 9046  
589 REPEAT 589 9066  
590 REPEAT 590 9086  
591 REPEAT 591 9106  
592 REPEAT 592 9126  
593 REPEAT 593 9146  
594 REPEAT 594 9166  
595 REPEAT 595 9186  
596 REPEAT 596 9206  
597 REPEAT 597 9226  
598 REPEAT 598 9246  
599 REPEAT 599 9266  
600 REPEAT 600 9286  
601 REPEAT 601 9306  
602 REPEAT 602 9326  
603 REPEAT 603 9346  
604 REPEAT 604 9366  
605 REPEAT 605 9386  
606 REPEAT 606 9406  
607 REPEAT 607 9426  
608 REPEAT 608 9446  
609 REPEAT 609 9466  
610 REPEAT 610 9486  
611 REPEAT 611 9506  
612 REPEAT 612 9526  
613 REPEAT 613 9546  
614 REPEAT 614 9566  
615 REPEAT 615 9586  
616 REPEAT 616 9606  
617 REPEAT 617 9626  
618 REPEAT 618 9646  
619 REPEAT 619 9666  
620 REPEAT 620 9686  
621 REPEAT 621 9706  
622 REPEAT 622 9726  
623 REPEAT 623 9746  
624 REPEAT 624 9766  
625 REPEAT 625 9786  
626 REPEAT 626 9806  
627 REPEAT 627 9826  
628 REPEAT 628 9846  
629 REPEAT 629 9866  
630 REPEAT 630 9886  
631 REPEAT 631 9906  
632 REPEAT 632 9926  
633 REPEAT 633 9946  
634 REPEAT 634 9966  
635 REPEAT 635 9986  
636 REPEAT 636 10006  
637 REPEAT 637 10026  
638 REPEAT 638 10046  
639 REPEAT 639 10066  
640 REPEAT 640 10086  
641 REPEAT 641 10106  
642 REPEAT 642 10126  
643 REPEAT 643 10146  
644 REPEAT 644 10166  
645 REPEAT 645 10186  
646 REPEAT 646 10206  
647 REPEAT 647 10226  
648 REPEAT 648 10246  
649 REPEAT 649 10266  
650 REPEAT 650 10286  
651 REPEAT 651 10306  
652 REPEAT 652 10326  
653 REPEAT 653 10346  
654 REPEAT 654 10366  
655 REPEAT 655 10386  
656 REPEAT 656 10406  
657 REPEAT 657 10426  
658 REPEAT 658 10446  
659 REPEAT 659 10466  
660 REPEAT 660 10486  
661 REPEAT 661 10506  
662 REPEAT 662 10526  
663 REPEAT 663 10546  
664 REPEAT 664 10566  
665 REPEAT 665 10586  
666 REPEAT 666 10606  
667 REPEAT 667 10626  
668 REPEAT 668 10646  
669 REPEAT 669 10666  
670 REPEAT 670 10686  
671 REPEAT 671 10706  
672 REPEAT 672 10726  
673 REPEAT 673 10746  
674 REPEAT 674 10766  
675 REPEAT 675 10786  
676 REPEAT 676 10806  
677 REPEAT 677 10826  
678 REPEAT 678 10846  
679 REPEAT 679 10866  
680 REPEAT 680 10886  
681 REPEAT 681 10906  
682 REPEAT 682 10926  
683 REPEAT 683 10946  
684 REPEAT 684 10966  
685 REPEAT 685 10986  
686 REPEAT 686 11006  
687 REPEAT 687 11026  
688 REPEAT 688 11046  
689 REPEAT 689 11066  
690 REPEAT 690 11086  
691 REPEAT 691 11106  
692 REPEAT 692 11126  
693 REPEAT 693 11146  
694 REPEAT 694 11166  
695 REPEAT 695 11186  
696 REPEAT 696 11206  
697 REPEAT 697 11226  
698 REPEAT 698 11246  
699 REPEAT 699 11266  
700 REPEAT 700 11286  
701 REPEAT 701 11306  
702 REPEAT 702 11326  
703 REPEAT 703 11346  
704 REPEAT 704 11366  
705 REPEAT 705 11386  
706 REPEAT 706 11406  
707 REPEAT 707 11426  
708 REPEAT 708 11446  
709 REPEAT 709 11466  
710 REPEAT 710 11486  
711 REPEAT 711 11506  
712 REPEAT 712 11526  
713 REPEAT 713 11546  
714 REPEAT 714 11566  
715 REPEAT 715 11586  
716 REPEAT 716 11606  
717 REPEAT 717 11626  
718 REPEAT 718 11646  
719 REPEAT 719 11666  
720 REPEAT 720 11686  
721 REPEAT 721 11706  
722 REPEAT 722 11726  
723 REPEAT 723 11746  
724 REPEAT 724 11766  
725 REPEAT 725 11786  
726 REPEAT 726 11806  
727 REPEAT 727 11826  
728 REPEAT 728 11846  
729 REPEAT 729 11866  
730 REPEAT 730 11886  
731 REPEAT 731 11906  
732 REPEAT 732 11926  
733 REPEAT 733 11946  
734 REPEAT 734 11966  
735 REPEAT 735 11986  
736 REPEAT 736 12006  
737 REPEAT 737 12026  
738 REPEAT 738 12046  
739 REPEAT 739 12066  
740 REPEAT 740 12086  
741 REPEAT 741 12106  
742 REPEAT 742 12126  
743 REPEAT 743 12146  
744 REPEAT 744 12166  
745 REPEAT 745 12186  
746 REPEAT 746 12206  
747 REPEAT 747 12226  
748 REPEAT 748 12246  
749 REPEAT 749 12266  
750 REPEAT 750 12286  
751 REPEAT 751 12306  
752 REPEAT 752 12326  
753 REPEAT 753 12346  
754 REPEAT 754 12366  
755 REPEAT 755 12386  
756 REPEAT 756 12406  
757 REPEAT 757 12426  
758 REPEAT 758 12446  
759 REPEAT 759 12466  
760 REPEAT 760 12486  
761 REPEAT 761 12506  
762 REPEAT 762 12526  
763 REPEAT 763 12546  
764 REPEAT 764 12566  
765 REPEAT 765 12586  
766 REPEAT 766 12606  
767 REPEAT 767 12626  
768 REPEAT 768 12646  
769 REPEAT 769 12666  
770 REPEAT 770 12686  
771 REPEAT 771 12706  
772 REPEAT 772 12726  
773 REPEAT 773 12746  
774 REPEAT 774 12766  
775 REPEAT 775 12786  
776 REPEAT 776 12806  
777 REPEAT 777 12826  
778 REPEAT 778 12846  
779 REPEAT 779 12866  
780 REPEAT 780 12886  
781 REPEAT 781 12906  
782 REPEAT 782 12926  
783 REPEAT 783 12946  
784 REPEAT 784 12966  
785 REPEAT 785 12986  
786 REPEAT 786 13006  
787 REPEAT 787 13026  
788 REPEAT 788 13046  
789 REPEAT 789 13066  
790 REPEAT 790 13086  
791 REPEAT 791 13106  
792 REPEAT 792 13126  
793 REPEAT 793 13146  
794 REPEAT 794 13166  
795 REPEAT 795 13186  
796 REPEAT 796 13206  
797 REPEAT 797 13226  
798 REPEAT 798 13246  
799 REPEAT 799 13266  
800 REPEAT 800 13286  
801 REPEAT 801 13306  
802 REPEAT 802 13326  
803 REPEAT 803 13346  
804 REPEAT 804 13366  
805 REPEAT 805 13386  
806 REPEAT 806 13406  
807 REPEAT 807 13426  
808 REPEAT 808 13446  
809 REPEAT 809 13466  
810 REPEAT 810 13486  
811 REPEAT 811 13506  
812 REPEAT 812 13526  
813 REPEAT 813 13546  
814 REPEAT 814 13566  
815 REPEAT 815 13586  
816 REPEAT 816 13606  
817 REPEAT 817 13626  
818 REPEAT 818 13646  
819 REPEAT 819 13666  
820 REPEAT 8

```

QY 1429 CCTCTGTTGAACCCAGCTTGAGACAGTGGAGCAGCAGCATGGTCTACCTGACACT 1488
Db 158 ProAlaThrSerPro-----ProGlyAsp 165
QY 1489 TCTGTGCTCCACCTGCTATGCGCTCTACCTCC-----CTGTCAAGAGCTCCACCTTTC 1542
Db 166 SerThrSerProAspHisSerThrSerProAlaThrArgAlaProGluAsp 185
QY 1543 TTTATGGCATCAAGCATCTCTCTGACTGAT-----CAA 1578
Db 186 SerThrSerThrAlaValLeuSerGlyThrSerSerProAlaThrThrAlaProValAsp 205
QY 1579 GGCACACACAGATCAATGCGCCACTGACACAGCACTAGTACACAGCGCTCCACCTCCCC 1638
Db 206 SerThrSerProValAlaHisAspThrSer---SerProAlaThrSerLeu--- 223
QY 1639 ACCAGTGATTATCTGCAATCAGCCACTGGCTCTGGGAATTTACATCCACTGCTCATCT 1698
Db 224 ---SerGluAspSerAlaSerSerProValAlaHisGlyGlyThrSerSerProAlaThr 242
QY 1699 TCAGAT-----GACAGCCGATCAAGTCCA----- 1722
Db 243 SerProLeuArgAspSerThrSerSerProValHisSerSerAlaSerIleGlnAsnIle 262
QY 1722 ----- 1722
Db 263 LysThrThrSerAspLeuAlaSerThrProAspHisAsnGlyThrSerValThrThrThr 282
QY 1723 -----GGTGGCGAAGATATGTCAGACAC 1746
Db 283 SerSerAlaLeuGlySerAlaThrSerProAspHisSerGlyThrSerThrThrAsn 302
QY 1747 CTAGATGAATGATCTGTCTGACACTCTGCCCCATCTGAGGTACACAGCTCAGCGAA 1806
Db 303 SerSerGluSerValLeuAlaThrProValTyrSerSerMetProPheSerThrThr 322
QY 1807 TATGTTTCT-----GTCCAGATCAT----- 1827
Db 323 LysValThrSerGlySerAlaIleIleProAspHisAsnGlySerSerValLeuProThr 342
QY 1828 -----TTCITGGAGATACCACT----- 1845
Db 343 SerSerValLeuGlySerAlaThrSerLeuValTyrAsnThrSerAlaIleAlaThrThr 362
QY 1846 -----CCTGCTCTCAGCTTTACAG 1863
Db 363 ProValSerAsnGlyThrGlnProSerValProSerGlnTyrProValSerProThrMet 382
QY 1864 TATATCACCACTAGTTCTATGACCATGCCCCAAGGCGGAGAGCTGTGTAGTGTCTTTC 1923
Db 383 AlaThrThrSerSerHisSerThrIleAlaSerSerTyrTyrSerThrValProPhe 402
QY 1924 AGT----- 1926
Db 403 SerThrPheSerSerAsnSerSerProGlnLeuSerValGlyValSerPhePheLeu 422
QY 1927 ---CTGCGTGTGCTACATGGCTTCTCCACGACCTGTTCAACAGAGCTCTCTGGAG 1983
Db 423 PhePheTyrIleGlnAsnHisProPheAsnSerSerLeuGluAspProSerSerAsnTyr 442
QY 1984 TACGAGCTCTGGAGCAACAATTCAACAGCTGTGTGTTCATATCTACGATCCAATCTT 2043
Db 443 TyrGlnGluLeuLysArgAsnIleSerGlyLeuPheLeuGlnIlePheAsnGlyAspPhe 462
QY 2044 ACAGGATTTTAAGCAACTTGAATTAATCTTCAAGAACGGGAGTGTGTTGTAATAGC 2103
Db 463 LeuGlyIleSerSerIleLys-----PheArgSerGlySerValValGluSer 479
QY 2104 AAAATCAAGTTTGTAACTGTGTGCGGTATAACCTCACCAAGGCTGTGACGGGCTTTC 2163
Db 480 ThrValValPheArgGluGlyThr-----PheSerAla 490
QY 2164 GAGGATTTTCGTTCTGTCGAGCCCAACACTCCATCTGGAAATAGACGACTACTCTCTC 2223

```

```

Db 491 SerAspValLysSerGlnLeuIleGlnHis---LysLysGluAlaAspSerTyrAsnLeu 509
QY 2224 AACATTGAA-----CCAGCTCATCAAGCAGATCCC 2253
Db 510 ThrIleSerGluValLysValAsnGluMetGlnPheProSerAlaGlnSerArgPro 529
QY 2254 -----TGCAAGTTCCTGGCTCGGC 2274
Db 530 GlyValProGlyTyrGlyIleAlaLeuLeuValLeuValCysIleLeuValAlaLeuAla 549
QY 2275 -----GAATTTGCCCAATGCTGTAAGAACGACGACTGAGGAAGCGGAG 2319
Db 550 IleValTyrPheLeuAlaLeuAlaValCys-----Gln 560
QY 2320 TGTGCGTGCACAAACAGGATATACAGCCAGGAGGAGCTCGAC 2361
Db 561 CysArgArgLys-----SerTyrGlyGlnLeuAsp 570

```

Search completed: March 4, 2004, 18:48:16  
Job time : 171.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 4, 2004, 18:38:06 ; Search time 161 Seconds  
(without alignments)  
13051.871 Million cell updates/sec

Title: US-10-007-270-1

Perfect score: 5875

Sequence: 1 taaccaagaaggtatcct.....tactatatgacataatcaat 3330

Scoring table: BLOSUM62

|             |             |
|-------------|-------------|
| Xgapop 10.0 | Xgapext 0.5 |
| Ygapop 10.0 | Ygapext 0.5 |
| Fgapop 6.0  | Fgapext 7.0 |
| Delop 6.0   | Delext 7.0  |

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

```
-MODEL=frame_n2p.model -DRV=xlp
-Q=/cgn2_1/USPTO_epool_p/US10007270/runat 04032004.160716.4957/app query.fasta_1.3527
-DB=SPTRMBL 25 -QFMT=fastan -SUFFIX=sept -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=opt -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10007270.qcgn_1_181 @runat 04032004.160716.4957 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database : SPTRMBL 25.\*

```
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvtrus.*
16: sp_bacteriap.*
17: sp_archaeap.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| 1          | 4073  | 69.3        | 797    | 4  | 043686 homo sapien |

|    | 2      | 2717 | 46.2  | 794 | 6      | Q9GMS5              |
|----|--------|------|-------|-----|--------|---------------------|
| 3  | 2504.5 | 42.6 | 798   | 11  | Q9ET31 | Q9GMS5 bos taurus   |
| 4  | 2504.5 | 42.6 | 798   | 11  | Q9ET62 | Q9ET31 mus musculus |
| 5  | 2502.5 | 42.6 | 798   | 11  | Q8RIW8 | Q9ET62 rattus norv  |
| 6  | 2343.5 | 39.9 | 742   | 11  | Q9S562 | Q8RIW8 mus musculus |
| 7  | 1810   | 30.8 | 928   | 13  | Q8JIR8 | Q9S562 mus musculus |
| 8  | 1450   | 24.7 | 288   | 4   | Q9BWZ1 | Q8JIR8 gallus gall  |
| 9  | 923    | 15.7 | 1241  | 4   | Q9UKK5 | Q9BWZ1 homo sapien  |
| 10 | 917    | 15.6 | 1241  | 4   | Q9UKD4 | Q9UKK5 homo sapien  |
| 11 | 916    | 15.6 | 1241  | 4   | Q9BZV3 | Q9UKD4 homo sapien  |
| 12 | 891    | 15.2 | 1127  | 11  | Q810Y3 | Q9BZV3 homo sapien  |
| 13 | 860    | 14.6 | 1239  | 11  | P70628 | Q810Y3 mus musculus |
| 14 | 855.5  | 14.6 | 1243  | 11  | Q80XH2 | P70628 rattus norv  |
| 15 | 841    | 14.3 | 185   | 6   | O46616 | Q80XH2 mus musculus |
| 16 | 664    | 11.3 | 215   | 11  | Q9CTP8 | O46616 macaca fasc  |
| 17 | 279    | 4.7  | 279   | 11  | Q8C8R4 | Q9CTP8 mus musculus |
| 18 | 196.5  | 3.3  | 16223 | 5   | Q8IR22 | Q8C8R4 mus musculus |
| 19 | 183.5  | 3.1  | 1831  | 10  | Q7KN11 | Q8IR22 drosophila   |
| 20 | 181    | 3.1  | 1305  | 10  | Q9FJ35 | Q7KN11 oryza sativ  |
| 21 | 165.5  | 2.8  | 631   | 11  | Q99K60 | Q9FJ35 arabidopsis  |
| 22 | 165    | 2.8  | 553   | 6   | Q9MZL1 | Q99K60 mus musculus |
| 23 | 164.5  | 2.8  | 1742  | 5   | Q810U0 | Q9MZL1 macaca mula  |
| 24 | 161    | 2.7  | 2011  | 3   | Q871L5 | Q810U0 plasmodium   |
| 25 | 158.5  | 2.7  | 1806  | 5   | Q869R4 | Q871L5 neurospora   |
| 26 | 157.5  | 2.7  | 1395  | 4   | Q96SB7 | Q869R4 dictyosteli  |
| 27 | 157.5  | 2.7  | 1759  | 5   | Q811M0 | Q96SB7 homo sapien  |
| 28 | 157    | 2.7  | 602   | 6   | O19115 | Q811M0 plasmodium   |
| 29 | 157    | 2.7  | 1621  | 16  | Q9KTA5 | O19115 oryctolagus  |
| 30 | 156    | 2.7  | 2922  | 5   | Q81PJ3 | Q9KTA5 vibrio chol  |
| 31 | 156    | 2.7  | 1749  | 5   | Q811L7 | Q81PJ3 drosophila   |
| 32 | 156    | 2.7  | 1758  | 5   | Q811L5 | Q811L7 plasmodium   |
| 33 | 155.5  | 2.6  | 768   | 16  | Q9KIK0 | Q811L5 plasmodium   |
| 34 | 155.5  | 2.6  | 960   | 4   | Q7Z4A6 | Q9KIK0 neisseria m  |
| 35 | 155    | 2.6  | 1763  | 5   | Q811L2 | Q7Z4A6 homo sapien  |
| 36 | 155    | 2.6  | 16215 | 5   | Q9NFS3 | Q811L2 plasmodium   |
| 37 | 154.5  | 2.6  | 1751  | 5   | Q810G9 | Q9NFS3 drosophila   |
| 38 | 154    | 2.6  | 1014  | 11  | Q8VWU7 | Q810G9 plasmodium   |
| 39 | 154    | 2.6  | 1030  | 11  | Q8C1C0 | Q8VWU7 mus musculus |
| 40 | 154    | 2.6  | 1433  | 16  | Q8DLX7 | Q8C1C0 mus musculus |
| 41 | 153.5  | 2.6  | 1728  | 5   | Q810M9 | Q8DLX7 synechococc  |
| 42 | 153    | 2.6  | 919   | 16  | Q9Z7K3 | Q810M9 plasmodium   |
| 43 | 153    | 2.6  | 1743  | 5   | Q811L9 | Q9Z7K3 chlamydia p  |
| 44 | 153    | 2.6  | 1802  | 5   | Q9VQU8 | Q811L9 plasmodium   |
| 45 | 153    | 2.6  | 17352 | 5   | Q95YV2 | Q9VQU8 drosophila   |
|    |        |      |       |     |        | Q95YV2 procambatus  |

#### ALIGNMENTS

RESULT 1  
043686 PRELIMINARY; PRT; 797 AA.  
AC 043686; Q95094;  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Interphotoreceptor matrix proteoglycan 150.  
DE INPGI OR IPML50.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Retina;  
RA MEDLINE=20068045; PubMed=10601738;  
RX Kuehn M.H., Hageman G.S.;  
RT "Expression and characterization of the IPM 150 gene (INPGI) product,  
RT a novel human photoreceptor cell-associated chondroitin-sulfate  
RL proteoglycan.";  
RL Matrix Biol. 18:509-518 (1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98358139; PubMed=9691169;

RA Felbor U., Gehrig A., Sauer C.G., Marquardt A., Kohler M., Schmid M.,  
 RT Weber B.H.;  
 RT "Genomic organization and chromosomal localization of the  
 RT interphosphoreceptor matrix proteoglycan-1 (IMP1) gene: a candidate for  
 RT 6q-linked retinopathies.";  
 RL Cytogenet. Cell Genet. 81:12-17(1998).  
 DR EMBL; AF047492; AAC03789.2; -;  
 DR EMBL; AF017776; AAC68835.1; JOINED.  
 DR EMBL; AF017760; AAC68835.1; JOINED.  
 DR EMBL; AF017761; AAC68835.1; JOINED.  
 DR EMBL; AF017762; AAC68835.1; JOINED.  
 DR EMBL; AF017763; AAC68835.1; JOINED.  
 DR EMBL; AF017764; AAC68835.1; JOINED.  
 DR EMBL; AF017765; AAC68835.1; JOINED.  
 DR EMBL; AF017766; AAC68835.1; JOINED.  
 DR EMBL; AF017767; AAC68835.1; JOINED.  
 DR EMBL; AF017768; AAC68835.1; JOINED.  
 DR EMBL; AF017769; AAC68835.1; JOINED.  
 DR EMBL; AF017770; AAC68835.1; JOINED.  
 DR EMBL; AF017771; AAC68835.1; JOINED.  
 DR EMBL; AF017772; AAC68835.1; JOINED.  
 DR EMBL; AF017773; AAC68835.1; JOINED.  
 DR EMBL; AF017774; AAC68835.1; JOINED.  
 DR EMBL; AF017775; AAC68835.1; JOINED.  
 DR Genew; HGNC:6055; IMP1.  
 DR GO; GO:0005578; C:extracellular matrix; TAS.  
 DR GO; GO:0005201; F:extracellular matrix structural constituent; TAS.  
 DR GO; GO:0007601; F:vision; TAS.  
 DR InterPro; IPR000082; SEA\_domain.  
 DR Pfam; PF01390; SEA; 2.  
 DR SMART; SM00200; SEA; 2.  
 DR PROSITE; PS50024; SEA; 1.  
 KW Receptor.  
 SQ SEQUENCE 797 AA; 89387 MW; D017ED090C78D521 CRC64;  
 Alignment Scores:  
 Pred. No.: 1.86e-289 Length: 797  
 Score: 4073.00 Matches: 796  
 Percent Similarity: 97.31% Conservative: 1  
 Best Local Similarity: 97.19% Mismatches: 0  
 Query Match: 69.33% Indels: 22  
 DB: 4 Gaps: 1  
 US-10-007-270-1 (1-3330) x 043686 (1-797)  
 QY 131 ATGATTTGGAACACTGAGAGAGCTATTTTGGATTTTTCCTCAAGTTCAAGGA 190  
 DB 1 MetTyrLeuGluThrArgAlaPheValPheTrpIlePheLeuGlnValGlnGly 20  
 QY 191 ACCAAGATATCTCCATTACATATACATACCTCTGAACTAAAGACATAGACATACCCCA 250  
 DB 21 ThrLysAspIleSerIleAsnIleTyrHisSerGluThrLysAspIleAspAsnProPro 40  
 QY 251 AGAATGAACACTGAAAGTACTGAAAAATGTACAAAAATGTCAACTATAGACGAATA 310  
 DB 41 ArgAsnGluThrGluSerThrGluLysMetTyrLysMetSerThrMetArgArgIle 60  
 QY 311 TTCCATTTCGCAAGCATCGACAAAGATCCGATTTTCCCAACGGGGGTAAAGTC 370  
 DB 61 PheAspLeuAlaLysHisArgThrLysArgSerAlaPhePheProThrGlyValLysVal 80  
 QY 371 TGTCACAGGAATCCATGAACAGATTTTAGACAGCTTCTCAAGCTTATATAGATTGAGA 430  
 DB 81 CysProGlnGluSerMetLysGlnIleLeuAspSerLeuGlnAlaTyrTyrArgLeuArg 100  
 QY 431 GTGTGTGAGAGAGATGATGGAGCATATCGATCTTTCTGGATCGCATCCCTGACACA 490  
 DB 101 ValCysGlnGlnAlaValTrpGluAlaTyrArgIlePheLeuAspArgIleProAspThr 120  
 QY 491 GGGGAATATCAGAGCTGGGTCCAGCATCTCCAGAGAGAGACCTTCTGGCTTTTGACATT 550  
 DB 121 GlyGluTyrGlnAspTrpValSerIleCysGlnGlnGlnThrPheCysLeuPheAspIle 140

QY 551 GGAATAAATCTCAGCAATTCCTCAGGAGCACTGGATCTTCTCCAGCAGAGATATAACAG 610  
 DB 141 GlyLysAsnPheSerAsnSerGlnGluHisLeuAspLeuGlnGlnArgIleLysGln 160  
 QY 611 AGAAGTTTCCCTGACAGAAAGATGAAATATCTCAGAGAAGACATTTGGGAGAGCTGGT 670  
 DB 161 ArgSerPheProAspArgLysAspGluLysSerAlaGluLysThrLeuGluGluProGly 180  
 QY 671 GAAACATTTGTATTTCAACAGCAATCTACATTTTCAAAGACTTGGGACAGTATTCTAAGAA 730  
 DB 181 GluThrIleValIleSer--Thr----- 187  
 QY 731 AACCTCAGAAGAGCAATTCAGATGTTGCCAAGCTTCTCACTTGGGCTTTTCCCTCTCA 790  
 DB 188 -----AspValAlaAsnValSerLeuGlyProPheProLeuT 200  
 QY 791 CTCTCATCAGACCTCTCTCAATGAAATCTCGATAATATACCTCAACGACACCAAGATGC 850  
 DB 200 hrProAspAspThrLeuLeuAsnGluLysLeuAspAsnThrLeuAsnAspThrLysMetP 220  
 QY 851 CTACACAGAGAAGAAACAGATTCGCTGTGTGGAGGAGCAGAGGTGGAGCTCAGCG 910  
 DB 220 roThrThrGluArgGluThrGluPheAlaValLeuGluGluGlnArgValGluLeuSerV 240  
 QY 911 TCTCTCTGTAAACAGAGTTCAAGCGCAGAGCTCCCTGACTCCACGTCCTCCATATTACC 970  
 DB 240 aIserLeuValAsnGlnLysPheLysAlaGluLeuAlaAspSerGlnSerProTyrTyG 260  
 QY 971 AGAGCTAGCAGGAAAGTCCCACTTCAGATGCAAAAGATATTTAAGAAATCTTCAGAGT 1030  
 DB 260 InGluLeuAlaGlyLysSerGlnLeuGlnMetGlnLysIlePheLysLeuProGlyP 280  
 QY 1031 TCAAAAAATCCATGCTTTAGATTTAGACCAAGAAAGAAAGATGGTCAAGTCCA 1090  
 DB 280 helLysIleHisValLeuGlyPheArgProLysGlyLysLeuAspGlySerSerSert 300  
 QY 1091 CAGAGATGCAACTTACGGCATCTTTAAGAGACACAGTCGACAGCAAGCAAAAGCCCTGAA 1150  
 DB 300 hrGluMetGlnLeuThrAlaIlePheLysArgHisSerAlaGluLysSerProAlaS 320  
 QY 1151 GTGACCTCTGCTTTTGATTCACAAAATTCGAAAGTGGAGAAAGTCTATCATGAAACA 1210  
 DB 320 erAspLeuLeuSerPheAspSerAsnLysIleGluSerGluGluValTyrHisGlyThrM 340  
 QY 1211 TGGAGGAGCAAGCAAGCAAGCAAGTATCTCTCAGAGCTACAGACCTCAAAAGCTGATCA 1270  
 DB 340 elGluGluAspLysGlnProGluLysThrLeuThrAlaThrAspLeuLysArgLeuLys 360  
 QY 1271 GCAAGCACTAGAGGAGAAACAATCTTTGGATGTGGGACAAATTCAGTTTCATGATGAAA 1330  
 DB 360 erLysAlaLeuGluGluGlnSerLeuAspValGlyThrIleGlnPheThrAspGluI 380  
 QY 1331 TTGCTGATCAGTCGACGCTTTGGTCTGTGACACCAATCAGAGCTGCCCATCTTTTG 1390  
 DB 380 leAlaGlySerLeuProAlaPheGlyProAspThrGlnSerGluLeuProThrSerPheA 400  
 QY 1391 CTGTATTAAACAGAGATCTACTTTTGAGTCCAGACTTCTCTCTGTTGAGACCCAGCTTG 1450  
 DB 400 laValIleThrGluAspAlaThrLeuSerProGluLeuProValGluProGlnLeuG 420  
 QY 1451 AGACAGTGGAGCGAGCAGCATGGTCTACCTGACACTTCTTGGTCTCCACCTGCTATGG 1510  
 DB 420 luThrValAspGlyAlaGluHisGlyLeuProAspThrSerTrpSerProAlaMetA 440  
 QY 1511 CCTTACTCTCTGTGAGAGCTCCACTTTCTTTTATGGCATCAAGCATCTTCTCTCTGA 1570  
 DB 440 laSerThrSerLeuSerGluAlaProProPhePheMetAlaSerSerIlePheSerLeuT 460  
 QY 1571 CTGATCAAGCAACCAAGATACATGCGCAGCTCCAGACAAATGCTTGTACAGGGCTCA 1630  
 DB 460 hrAspGlnGlyThrThrAspThrMetAlaThrAspGlnThrMetLeuValProGlyLeuT 480  
 QY 1631 CCATCCCAACGATGATTATTCTGCAATCAGCAACTGGCTCTGGGAATTTTACATCCAC 1690



```

Db      480 hrllleProThrSerAspTyr-SerAlaIleSerGlnLeuAlaLeuGlyIleSerHisProp 500
QY      1691 CTGCATCTTCAGATCAGACGCGATCAAGTCAGGTGCGAAGATATGTCAGACACTAG 1750
Db      500 roalaSerSerAspSerArgSerSerAlaGlyGlyGluAspMetValArgHisLeuA 520
QY      1751 ATGAATGGATCTGTGACACTCTCGCCCATCTGAGGTACAGAGTCCAGGAGTATG 1810
Db      520 spGluMetAspLeuSerAspThrProAlaProSerGluValProGluLeuSerGluTyrV 540
QY      1811 TTTCTCTCCAGATCATTTCTTGAGGATACACTCTCTGTCTCAGCTTTACAGTATATCA 1870
Db      540 alSerValProAspHisPheLeuGluAspThrThrProValSerAlaLeuGlnTyrIle 560
QY      1871 CCCTAGTCTATCAGCATTCGCCCCAGGCGGAGAGTGTAGTGTCTTCTCAGTCTGC 1930
Db      560 hrThrSerSerMetThrIleAlaProLysGlyArgGluLeuValValPhePheSerLeuA 580
QY      1931 GTTGTCTTAACATGGCCTTCTTCAACGACCTGTTCACAGAGCTCTCTGAGTACCGAG 1990
Db      580 rgValAlaAsnMetAlaPheSerAsnAspLeuPheAsnLysSerSerLeuGluTyrArgA 600
QY      1991 CTCTGAGCAACATTCACACAGCTGCTGTTCCATATCTACGATCCAACTTTACAGAT 2050
Db      600 laeGluGlnGlnPheThrGlnLeuLeuValProTyrLeuArgSerAsnLeuThrGlyP 620
QY      2051 TTAAGCACTTGAATACTTAATCTTCAAGACGGAGTGTGATTCGAATAGCAAAATGA 2110
Db      620 heLysGlnLeuGluIleLeuAsnPheArgAsnGlySerValIleValAsnSerLysMetL 640
QY      2111 AGTTTCTAGTCTGTGCGGTATTAACCTCAAGCTGTGACGGGCTCTTGGAGGATT 2170
Db      640 yePheAlaLysSerValProTyrAsnLeuThrLysAlaValHisGlyValLeuGluAsp 660
QY      2171 TTCTGCTGTGTCAGCCCAACACTCCATCTGGAATAGACAGTACTCTCTCAACATTG 2230
Db      660 heArgSerAlaAlaGlnGlnLeuHisLeuGluIleAspSerTyrSerLeuAsnIleG 680
QY      2231 AACAGCTGATCAAGCAGATCCCTGCAAGTCTCTGCGCTGCGGCGAATTTGCCATGTG 2290
Db      680 luProAlaAspGlnAlaAspProCysLysPheLeuAlaCysGlyGluPheAlaGlnCysV 700
QY      2291 TAAGAAGCAACGGACTGAGGACCGGAGTGTGCTGCAACACAGGATATGACAGCCAGG 2350
Db      700 alLysAsnGluArgThrGluGluAlaGluCysArgCysLysProGlyTyrAspSerGln 720
QY      2351 GGAGCTGGACGCTCTGGAACCGCTCTGTGGCTGCGCCCTGGCAAAAGGATGCGAGTCC 2410
Db      720 lySerLeuAspGlyLeuGluProGlyLeuCysGlyProGlyThrLysGluCysGluVal 740
QY      2411 TCCAGGAAGAGGAGCTCCATGACAGTTCAGATCACTCTGAAATCAAGCATACAA 2470
Db      740 euGlnGlyLysGlyAlaProCysArgLeuProAspHisSerGluAsnGlnAlaTyrIle 760
QY      2471 CTAGTGTAAAAGTTCAAAATCAAAATPAACAGGTAATCAGTAAAGAAATTCG 2530
Db      760 hrSerValLysLysPheGlnAsnGlnAsnLysValIleSerLysArgAsnSerG 780
QY      2531 AATTACTGACCTGAGATATGAGATTTTAACCATCAAGATTGGGAGGAAAT 2583
Db      780 luLeuLeuThrValGluTyrGluGluPheAsnHisGlnAspTrpGluGlyAsn 797

```

## RESULT 2

```

Q9GMS5
ID Q9GMS5 PRELIMINARY; PRT; 794 AA.
AC Q9GMS5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE MLGAPC core protein.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
RN NCBI_TaxID=9913;
RP SEQUENCE FROM N.A.
EX MEDLINE=20391374; PubMed=10937595;
RA Uehara F., Ooba N., Ozawa M.;
RT "Isolation and Characterization of Mucinlike Glycoprotein Associated
RT with Photoreceptor Cells.";
RL Invest. Ophthalmol. Vis. Sci. 41:2759-2765(2000).
DR EMBL; AB047844; BAB1254.1; -.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 2.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00200; SEA; 2.
DR PROSITE; PS00024; SEA; 1.
DR PROSITE; PS00024; SEA; 1.
SQ SEQUENCE 794 AA; 89377 MW; 4965FD322FA95841 CRC64;

```

## Alignment Scores:

```

Pred. No.: 3.54e-190 Length: 794
Score: 2717.00 Matches: 558
Percent Similarity: 76.90% Conservative: 81
Best Local Similarity: 67.15% Mismatches: 142
Query Match: 46.25% Indels: 51
DB: 6 Gaps: 8

```

US-10-007-270-1 (1-3330) x Q9GMS5 (1-794)

```

QY 131 ATGATTTGGAACTAGACAGCTATTTTGTGTTTGGATTTTCTCCAGTTCAGGA 190
Db 1 MetHisLeuGluAlaAlaArgValIlePhePheLeuTyrIlePheLeuGlnValGlnGly 20
QY 191 ACCAAGATATCTCCATTAACATATACCATCTCGAACTTGAACCTAAAGACATAGCAATCCCCA 250
Db 21 IleLysAspLeuSerIleLysIleTyrGlySerGluIleLysAspIleAspAsnAlaPro 40
QY 251 AGAATTAAGCAACTGAAATGTAACAAATGTAACAAATGTAACAAATGTAACAAATGTAACAAATG 310
Db 41 ArgThrGluAlaThrLysAsnThrAlaLysThrThrLysValSerThrMetArgArgIle 60
QY 311 TTGATTTGGCAAGCATCGAACAAGATCCGCAATTTTCCCAAGGGGGTTAAAGTC 370
Db 61 PheAspLeuAlaLysHisArgThrLysArgSerAlaPhePheProThrGlyValLysVal 80
QY 371 TGTCCACAGGAATCCATGAACAGATTTTACAGCTCTTCAAGCTTATATAGATTGAGA 430
Db 81 CysProGlnGluSerMetGluIleLeuAlaSerLeuGlnAlaTyrTyrArgLeuArg 100
QY 431 GTGTCTCAGGAACAGTATGGGAAGCATATCGGATCTTCTGGATCGCATCCCTGCACACA 490
Db 101 ValCysGlnGluAlaValTyrGluAlaTyrArgIlePheLeuAspArgLeuProGluPro 120
QY 491 GGGGAATATCAGGACTGGCTGAGCTCTGCGAGTCTGCGAGGAGACCTTCTGCTCTTTGACATT 550
Db 121 GlyGluTyrGlnAspTyrValSerValCysGlnGlnGlnThrPheCysLeuPheAspIle 140
QY 551 GGAAAAAATCTCAGCAATTTCCAGAGCACCTGGATCTTCCAGCAGAGATATAAAGCAG 610
Db 141 GlyLysAsnPheSerAsnSerGlnIleHisLeuAspLeuLeuGlnGlnArgMetLysGln 160
QY 611 AGAATTTCCCTGACAGAAAAGATGAATATCTCGACAGAGAAGACATTTGGAGAGCCTGGT 670
Db 161 ArgAsnPheLeuGluArgLysAspGluValValThrLysGluThrLeuGlyGluLeuGly 180
QY 671 GAAACCATGTTTCATTTCAACAGCAATCTACATTTTCAAAAGCTTGGGCGAGTATTTCAAGAA 730
Db 181 GlnThrProGlyLeuGln----- 186
QY 731 AACCTCAGAGAGCAAAATCAAGATGTTGCCAACGTCCTACCTTGGGCTTTTCCCTCTCA 790
Db 187 -----GlnThrLeuProValSerHisProGlyProCysLeuSer 199

```

```

QY 791 CTCTGATGACACCTCTCTCAATGAAATTCGTAATACATCTCAACGACACCAAGATG- 849
Db 200 LeuProMetThrAlaGlnArgAenProGlnLeuHisProSerArgThrProArgVal 219
QY 850 CCTACACAAAGAGAGAAACAGAAATTCGCTGTGTGAG-----GAGCAGAGG 897
Db 220 ProThrArgGluArgLysIleGluPheThrAspAlaAlaGluAspAlaLeuGluGlnLys 239
QY 898 GTGAGCTCAGGCTCTCTCTGTTAAACACAGAGTTCAGGCAGAGCTCGTCACTCCAG 957
Db 240 ValGluLeuSerIleSerLeuAlaAenGlnLysPheLysSerGluLeuAspAenSerGln 259
QY 958 TCCCAATATTACAGAGCTAGCAGAGAGTCCCAAGTCCAGATTCAGATCAAAAGATATTAA 1017
Db 260 SerProTyrTyrLeuGluValAlaAlaLysSerGlnLeuGlnMetGlnLysIlePheLys 279
QY 1018 AATCTTCAGGATCAAAATAATCCATGTGTAGGATTTAGACCAAGAAAGAAAAGAT 1077
Db 280 LysLeuProGlyPheLysGluIleHisValSerGlyPheArgProLysLysGluArgAsp 299
QY 1078 GGCTCAAGCTCCAGAGATGCACTTACGGCCATCTTTAAGACACAGTGCAGAGCA 1137
Db 300 GlyThrSerSerThrGluMetGlnLeuThrAlaIlePheLysLysGlyLysAlaGluAla 319
QY 1138 AAAAGCCCTCAAGTACCTCTCTCTTTCATTCACAAATAATGAAAGTGGAGAGTC 1197
Db 320 LysSerProAlaSerAspProLeuSerLeuAspSerAsnLysIleGluSerGluGlyAsp 339
QY 1198 TATCATGACATGGAGGAGGACCAAGCAACCAAGAAATCTATCTCACAGCTACAGACTC 1257
Db 340 ProArgGlyThrGluGluGluLysGluArgGluLeuTyrProThrAlaSerGluLeu 359
QY 1258 AAAAGCTGATCACAAGCACTAGAGAGAGAACATCTTGGATGTGGGACAAATTCAG 1317
Db 360 ArgLysLeuIleSerArgAlaLeuGluGluAspGlnSerLeuAspValGlyThrIleGln 379
QY 1318 TTCACGTGAAATGTCTGATCACTGCCAGCTCTTGGTCTGTGACCCCAATCAGAGCTG 1377
Db 380 PheThrAspGluIleValGlySerLeuProSerLeuAspProAspThrGlnLeuValLeu 399
QY 1378 CCCACATCTTTGCTGTTATACAGAGATGCTACTTTGAGTCCAGAACTTCTCTCTCT 1437
Db 400 ProThrLeuLeuThrAspIleThrLysAspAlaThrLeuSerProGluLeuProLeuGly 419
QY 1438 GAACCCAGCTTCAGACAGTGCAGCGAGCAGACATGCTACTGACACTTCT----- 1491
Db 420 GlnProArgLeuGluThrValAspArgAlaGlyHisSerProProGlyAlaSerProThr 439
QY 1492 -----TGGTCTCCACCTCTATGGCTCTACCTCCCTGTCAGAGCTCCACTTCTTT 1545
Db 440 AspGlyTyrSerProAlaMetThrSerThrSerLeuSerGluThrLeuProPhePhe 459
QY 1546 ATGGCATCAGCATCTCTCTGATCATCAAGGACACACAGATACAAATGCCCACTGAC 1605
Db 460 ThrAlaSerSerValPheProGlnThrAspGlnSerAlaThrAspIleMetSerIleAsp 479
QY 1606 CAGCAATGCTAGTACAGGCTCACCATCCACAGTATTTCTGATGATTTCTCAATCAGCAA 1665
Db 480 GlnThrValLeuLeuProArgLeuThrValProThrAspAspTyrSerAlaIleSerPro 499
QY 1666 CTGGCTTCGGAAATTCATCCACCTCATCTTCAGATGACAGCGATCAAGTGCAGGT 1725
Db 500 LeuValProGluIleSerHisLeuProThrSerSerGluAspThrLeuSerSerSer--- 518
QY 1726 GCGCAATATGTCAGACACTAGATGAATGGATCTGTCGACACTCTCTGCCCATCT 1785
Db 519 SerGlnAspThrMetLysLeuAspGlyValAspLeuThrLysThrProThrSerSer 538
QY 1786 GAGGTACAGAGCTCAGC-----GAATATGTTCTCTCCAGATCATTTCC 1830
Db 539 GluGlyProArgAsnSerValGlyMetPheProAlaThrIleIle-----Phe 554
QY 1831 TTGAGGAGTACCATCTCTGCTCAGCTTTTACAGTATATCACCAGTATGTTCTATGACCAT 1890

```

```

Db 555 LeuGluAsnIleThrProAspProGlyLeuArgTyrIleThrThrSerAlaMetThrVal 574
QY 1891 GCCCCCAGGCGGAGAGCTGCTAGTGTCTTCAGTCTGCGTGTCTTCAATGCGCTTC 1950
Db 575 AlaAlaArgGlyArgGluLeuValPhePheSerLeuArgValAlaAenValProPhe 594
QY 1951 TCCAAACACCTGTCACACAGAGCTCTCGAGTAGTACGAGCTCTGAGCAGCAATTCACA 2010
Db 595 SerThrAspLeuPheAsnLysSerSerLeuGluTyrGlnAlaLeuGluGlnArgPheThr 614
QY 2011 CAGCTGCTGCTCCATATCTACATCCAATCTTACAGATTTAAGCACTTGAATAACTT 2070
Db 615 GlnLeuLeuValProAsnLeuArgSerAsnLeuThrGlyPheLysGlnLeuGluIleLeu 634
QY 2071 AACTTCAGAAACGGAGTGTGATTTGAATAGCAAAATGAAGTTTGTCTGTAAGTACTT 2130
Db 635 AsnPheArgAsnGlySerValIleValAsnSerLysValArgPheAlaLysSerValPro 654
QY 2131 TATAACTCTCAACAGGCTGTGCACGGGTCTTGGAGGATTTTCTGCTGCGAGCCAA 2190
Db 655 TyrAsnLeuThrLysAlaValArgGlyValLeuGluAspPheArgSerThrAlaAlaGln 674
QY 2191 CAATCCATCTGGAATAGACAGCTACTCTCTCAACATTTGAACAGCTGATCAAGCAGAT 2250
Db 675 GlnLeuAspLeuGluIleAspSerTyrSerLeuAspValGluProAlaAspGlnAlaAsp 694
QY 2251 CCTGTCAGTCTTCCGCTCGCGCAATTTGCCCATGTGTAAAGAACAGAGCTGAG 2310
Db 695 ProCysLysPheLeuAlaCysGlyGluPheAlaGlnCysValArgAsnGluThrThrGlu 714
QY 2311 GAACGGAGTGTGCTGCAACACAGATATCAGACCCAGGAGGCTCGACGCTCTGGA 2370
Db 715 GluAlaGluCysArgCysArgSerGly-----ThrGlnAlaLeuValLeuProIleGlu 732
QY 2371 CAGGCTCTGTGCGCTCGCCCAAGAAATGCGAGTCTCTCCAGGGAAGGAGCTCCA 2430
Db 733 -----AspCysGluAspIleProGlyLysGlyThrPro 743
QY 2431 TGCAGTTGCCAGATCACTCTGAAATCAAGCATACAAACATGATGTTAAAGATTCOA 2490
Db 744 CysArgSerLeuAspGlnSerLysAsnGlnValTyrGluProGlyValLysPheGln 763
QY 2491 AATCAACAAAATAACAGGTATCAGTAAAGAAATTCGTAATTACTGACCGTAGAATAT 2550
Db 764 ArgGlnGlnAspAsnLysValThrMetLysArgLysPheGluLeuLeuThrIleGlyTyr 783
QY 2551 GAAGAAATTAACCATCAAGATTCGGAAGGAAAT 2583
Db 784 GluGluPheAsnTyrGlnAspTrpGluGlyAsn 794

RESULT 3
Q9ET31 PRELIMINARY; PRT; 798 AA.
AC Q9ET31;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE IMP 150 proteoglycan.
GN IMPG1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuehn M.H., Wietcki D.T., Hageman G.S.;
RT "Molecular characterization of the murine orthologue of the human
RL retinal proteoglycan IMP 150.";
RL Mol. Vis. 6:0-0(2000)
DR EMBL; AF266478; AAC00796.1; -.
DR MGD; MGI:1926876; Impg1.
DR InterPro; IPR000082; SEA_domain.

```

DR Pfam; PF01390; SEA; 2.  
DR SMART; SMO0200; SEA; 2.  
SQ SEQUENCE 798 AA; 89542 MW; 170D2CE94ACD76A2 CRC64;

Alignment Scores:

|                        |           |               |    |
|------------------------|-----------|---------------|----|
| Pred. No.:             | 1.28e-174 | Length:       | 79 |
| Score:                 | 2504.50   | Matches:      | 52 |
| Percent Similarity:    | 73.13%    | Conservative: | 80 |
| Best Local Similarity: | 63.49%    | Mismatches:   | 18 |
| Query Match:           | 42.63%    | Indels:       | 43 |
| DB:                    | 11        | Gaps:         | 10 |

US-10-007-270-1 (1-3330) x Q9ET31 (1-798)

|  |    |     |  |      |
|--|----|-----|--|------|
|  | QY | 131 | ATGTAATTGGAACTAAGACAGCTAATTTTGTGTTTTTGATTTTCCTCCAGTCAAGGA          | 190  |
|  | Db |     | :    :    :    :    :    :    :    :    :    :    :    :    :    : |      |
|  |    | 1   | MecAsnPheGlnIleuysHisalailePheValPheGlyllePheIeuGlnValGlnsly       | 20   |
|  | QY | 191 | ACCAAAGATATCTCCATTAACATATACCATCTGAAACTAAAGACATAGACAATCCGCCA        | 250  |
|  | Db |     | :    :    :    :    :    :    :    :    :    :    :    :    :    : |      |
|  |    | 21  | IleLysAspThrSerlleLysillePheSerSerGlulleLysasnilleasplysThrPro     | 40   |
|  | QY | 251 | AGAATAGAAACAACCTGAAGTACTGAAAATAATGTACAAATGTCAATAGACAGCAATA         | 310  |
|  | Db |     | :    :    :    :    :    :    :    :    :    :    :    :    :    : |      |
|  |    | 41  | ArglleGluThrleGluSerThrSerThrValHisLysValSerThrMetLysArgille       | 60   |
|  | QY | 311 | TTCGATTTGGCAAGCATCGAACAAAAAGATCCGATATTTTCCCAACGGGGTTAANGTC         | 370  |
|  | Db |     | :    :    :    :    :    :    :    :    :    :    :    :    :    : |      |
|  |    | 61  | PheAspLeuProLysLeuArgThrLyAsqSeraLaLeuPhePro--AlaLaAsnlie          | 79   |
|  | QY | 371 | TCGCCACAGNATCCATGAACACAGATTTTAGACATCTTCAAGCTTATTATAGATTGAGA        | 430  |
|  | Db |     | :    :    :    :    :    :    :    :    :    :    :    :    :    : |      |
|  |    | 80  | CysProGlnGluSerLeuArgGlnleLeuAlaSerLeuGlnGluTyTyArgLeuArg          | 99   |
|  | QY | 431 | GNGTCTCAGGAACGACTGTGGGAAGCATATCCGATCTTCTGCATGCATCCCTGACACA         | 490  |
|  | Db |     | :    :    :    :    :    :    :    :    :    :    :    :    :    : |      |
|  |    | 100 | Val.CysGlnGluValValTrpGluIatArYArgillePheLeuAspArgilleProAspThr    | 119  |
|  | QY | 491 | GGGGAATATCAGACTGGGTGAGCATCTGCCACGACGAGAGACCTTCTGCCTCTTTGCATT       | 550  |
|  | Db |     | :    :    :    :    :    :    :    :    :    :    :    :    :    : |      |
|  |    | 120 | GluGluTyrglnAspTrpValserLeuCysGlnlysGluThrPheCysIeupHeAspille      | 139  |
|  | QY | 551 | GGMAAAACTTCAGCANATCCCAGAGCACCTGGATCTTCTCCAGCAGAGAATAAAACAG         | 610  |
|  | Db |     | :    :    :    :    :    :    :    :    :    :    :    :    :    : |      |
|  |    | 140 | GlyLysAsnPheSerAsnSerGlnGluHisteuAspLeuleuGlnGlnArgilleLysGln      | 159  |
|  | QY | 611 | ACAGATTTCCTCGACAGAAAAGATGAATATCTGCACAGAGACATTTGGGAGAGCTCGT         | 670  |
|  | Db |     | :    :    :    :    :    :    :    :    :    :    :    :    :    : |      |
|  |    | 160 | ArgSerPheProGlyArgLysAspGluThrAlaSerMetGluThrLeuGluAlaProThr       | 179  |
|  | QY | 671 | GAACATTGTCAATTCACAGCAATCTACATTTCAAAGACTTGGGAGTAGTATTCAGNA          | 730  |
|  | Db |     | :    :    :    :    :    :    :    :    :    :    :    :    :    : |      |
|  |    | 180 | GluAlaProValpro-Thr-----186  |      |
|  | QY | 731 | AACCTCAGACAGCAAMATCAAGATGTGGCAACGTCACTTCTGGGCGCTTCCCTCTCA          | 790  |
|  | Db |     | :    :    :    :    :    :    :    :    :    :    :    :    :    : |      |
|  |    | 187 | -----AspValSerArgMetSerLeuGlyPropheProLeup                         | 199  |
|  | QY | 791 | CTCCTGATGACACCTCTCCAATCAATCTCGATAATACACTCAACGACACCAAGATGC          | 850  |
|  | Db |     | :    :    :    :    :    :    :    :    :    :    :    :    :    : |      |
|  |    | 199 | roSerAspAspThrAspIeumylgIleLeuSerValThrLeuLysAspilleGlnLysp        | 219  |
|  | QY | 851 | CTACAACAGAAAGAGAACA-----GAAATTGCTGTGTGGAGGAGC                      | 892  |
|  | Db |     | :    :    :    :    :    :    :    :    :    :    :    :    :    : |      |
|  |    | 219 | roThrThrGluSerlysThrGluProIleHisValserGluPheSer-----SerGluG        | 237  |
|  | QY | 893 | AGAGGTGAGAGCTCAGGCTCTCTTGTAACACCAGAAGTTCAAGCAGAGCTCGCTGACT         | 952  |
|  | Db |     | :    :    :    :    :    :    :    :    :    :    :    :    :    : |      |
|  |    | 237 | luLysValGluPheSexIleSerLeuProAsnHisargPhelysAlaGluLeuThrAnsS       | 257  |
|  | QY | 953 | CCAGTCCCCCATATTACCAGAGCTAGACAGAAAGTCCCACTTCAGATGCAAAAGATAT         | 1012 |
|  | Db |     | :    :    :    :    :    :    :    :    :    :    :    :    :    : |      |
|  |    | 257 | erGlySerProTyrryrdngInluEuvalGlyGlnSerGlnleuGlnleuGlnLysleip       | 277  |

|    |      |   |      |
|----|------|---|------|
| QY | 1013 | TTAAGAAATCTCCAGGATTCAAAATAATCCATGTTAGAGATTTAGACAAAGAGAA       | 1072 |
| Db |      |   |      |
| Db | 277  | heliysylsLeuProGlyPheGlyGluileargValLeuGlyPheArgProLysylsGluG | 297  |
| QY | 1073 | AAGATGGCTCAAGCTCCACAGAGATGCACACTTACGGCCATCTTTAAGAGACACAGTCGAG | 1132 |
| Db |      |   |      |
| Db | 297  | luaspGlySerSerThrGluilegLeuLeuMetAlailePheLysArgspHisAlaG     | 317  |
| QY | 1133 | AAGCAAAAAGCCCTGCAAGTGACCTCTGCTGCTTTTGTATCCAAACAAATGGAAGTCAGG  | 1192 |
| Db |      |   |      |
| Db | 317  | luAlaLysSerProaspSerHisLeuLeuSerLeuaspSerAsnLysileGluSerGluA  | 337  |
| QY | 1193 | AGTCTATATPAGAACCATGGAGGAGGACAGCAACACAGAAATCTATCTCACAGCTACAG   | 1252 |
| Db |      |   |      |
| Db | 337  | rgileHisGlyValile--GluaspLysGlnProGluThrTyrLeuthrAlaThrA      | 356  |
| QY | 1253 | ACCTCARAAGCGTCATCAGCAAAACGCTAGAGAAAGAAACAATCTTTGGATGTGGGGCAA  | 1312 |
| Db |      |   |      |
| Db | 356  | spLeuLysLysLeuileilegLeuLeuLeuaspGlyAspLeuSerLeuValGluGlyLysI | 376  |
| QY | 1313 | TTCACTTCACTGATGAATTTGCTGATCACTCCAGCCTTGTGCTGACAGCCCAATCAG     | 1372 |
| Db |      |   |      |
| Db | 376  | leProPheGlyAspGluValThrGlyThrLeu-----PheArgProValThrGluProA   | 394  |
| QY | 1373 | AGCTGCCACATCTTTTGTCTTATAACAGAGATGCTACTTTGACTGCAGAACTCTCCCTC   | 1432 |
| Db |      |   |      |
| Db | 394  | spLeuProLysProLeuAlaAspValThrGluaspAlaThrLeuSerProGluLeuProP  | 414  |
| QY | 1433 | CTGTGTAACCCAGCTTCAGACAGCTGGAGCGAGCMAGCATGCTTACCT-----1482     |      |
| Db |      |   |      |
| Db | 414  | heValGluProArgLeuGluAlaValAspArgGluGlySerGluLeuProGlyMetSerS  | 434  |
| QY | 1483 | -----GACACTTCTGGTCTCCACTGCTATGGCTCTACTCTCCCTGTCAGAGCTCCAC     | 1537 |
| Db |      |   |      |
| Db | 434  | erLysAspSerSerIrpSerProProValSerAlaSerIleSerArgSerGluAsnLeuP  | 454  |
| QY | 1538 | CTTCTTTATGGCATCAAGCATCTTCTCTGACTGATCAAGGCCACACAGATACATGCG     | 1597 |
| Db |      |   |      |
| Db | 454  | roSerPhe---ThrProSerIlePheSerLeuaspAlaGlnSerProProLeuMetI     | 473  |
| QY | 1598 | CCACTGCACAGACATGCTAGTACAGGGCTCACATCCCAACCAAGTATTATTCGCAA      | 1657 |
| Db |      |   |      |
| Db | 473  | hrThrGlyProThrAlaLeuIleProLysProThrLeuProThrIleAspTyrSerThrI  | 493  |
| QY | 1658 | TCAGCCAACTGCGCTCTGGGAATTTACATCCACCTGCATCTTCAGATCAGACGCGCATCA  | 1717 |
| Db |      |   |      |
| Db | 493  | leArgGlnLeuProLeuGluSerSerHisIrpProAlaSerSerAspArgGluLeuI     | 513  |
| QY | 1718 | GTGCAAGTGGCGAGATATGGTCAGACACCTAGATGAAATGGATCTGTCTGACACTCTCG   | 1777 |
| Db |      |   |      |
| Db | 513  | leThrSerSerHisAspThrIleArgAspLeuaspGlyMetAspValSerAspThrProA  | 533  |
| QY | 1778 | CCCCATCTCAGGTACCGAGCTTCGCGAATATGTTCTCTCCAGATCATTTCTTGGAGG     | 1837 |
| Db |      |   |      |
| Db | 533  | laLeuSerGluIleSerGluLeuSerGlyTyrAspSerAlaSerGlyGlnPheLeuGluW  | 553  |
| QY | 1838 | ATACCACTCTCTGCTTCAGCTTTACAGTATATACCACTAGTTCTATGACCATTTGCCCA   | 1897 |
| Db |      |   |      |
| Db | 553  | etThrThrProIleProThrValArgPheIleThrThrSerSerGluThrIleAlaThrL  | 573  |
| QY | 1898 | AGGGCCGAGAGCTGGTAGTGTTCATCTGCTGGTGTGTGCTAAATGGCTTCTCCAACG     | 1957 |
| Db |      |   |      |
| Db | 573  | y8GlyGlnIleValValaPhePheSerLeuArgValAlaAsnMetProPheSerTyrA    | 593  |
| QY | 1958 | ACCTGTTCAACAGAGCTCTCTGGAGTACCGAGCTCTGGAGCAACAATTCACAGAGTGC    | 2017 |
| Db |      |   |      |
| Db | 593  | spLeuPheAsnLysSerSerLeuGluTyrGlnAlaLeuGluGlnArgPheThrAspLeuL  | 613  |
| QY | 2018 | TGTTTCCATATCTACGATCCAATCTTACAGATTTTAAGCACTTGAATTACTTACTTCA    | 2077 |
| Db |      |   |      |
| Db | 613  | euValProTyrLeuArgSerAsnLeuThrGlyPheLysGlnLeuGluIleLeuSerPheA  | 633  |

```

QY 2078 GAAAGGGAGTGTGATTCTGAATAGCAAAATCAAGATTTCCTAAGTCTGTGCGGCTATAC 2137
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
633 rGAnGlySerValIleValAsnSerLysValArgPheAlaLysAlaValProTyrAsnL 653
QY 2138 TCACCAAGCTGTGCACGGGGTCTTGGAGGATTTCTGCTGTCAGCCCAACCACTCC 2197
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
653 eutHrGlnAlaValArgGlyValIleGluAspLeuArgSerThrAlaAlaGlnGlyLeuA 673
QY 2198 ATCTGGAATACAGACTACTCTCTCAATGAACCAAGCTGATCAAGCAGATCCCTGCA 2257
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
673 snLeuGluIleGluSerTyrSerLeuAspIleGluProAlaAspGlnAlaAspProCysL 693
QY 2258 AGTTCTCTGCGCTGCGCGCAATTTGCCCAATGTGTAAAGACGACGAGCTGAGAGCGG 2317
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
693 ySLeuLeuAspCysGlyIlePheAlaGlnCysValIleAsnGluTrpThrGluGluAlaG 713
QY 2318 AGTGTCTGCGTCAAAACAGGATATGACAGCCAGGGAGCTGACCGTCTGGAACCAAGCC 2377
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
713 luCysArgCysArgGlnGlyHisGluSerHisGlyThrLeuAspTyrGlnThrLeuAsnL 733
QY 2378 TCTGTGCGCTCGCACAAAGGAATCGGAGTCTCCAGGGAAAGGGAGCTCCATGCGAGT 2437
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
733 euCysProProGly---LysThrCysValAlaGlyArgGluGlnAlaThrProCysArgp 752
QY 2438 TGCAGATCACTCTGAAATCAAGCATACAAACTAGTGTAAAGTCTCAAAATCAAC 2497
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
752 roProAspHisSerThrAsnGlnAlaGlnGluProGlyValIleLysLeu-----Argg 770
QY 2498 AAATACAGAGTATCAGTAAAGAAATCTGAAATTAAGTACGCTAGATATGAAGAT 2557
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
770 InGlnAsnLysValValLysLysArgAsnSerLysLeuSerAlaIleGlyPheGluGluP 790
QY 2558 TTAACCATCAAGATTGGAGCAAT 2583
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
790 heGluAspGlnAspTrpGluGlyAsn 798

```

## RESULT 4

```

Q9ET62 ID Q9ET62 PRELIMINARY; PRT; 798 AA.
AC Q9ET62;
DT 01-MAR-2001 (TremBurel. 16, Created)
DT 01-MAR-2001 (TremBurel. 16, Last sequence update)
DT 01-JUN-2002 (TremBurel. 21, Last annotation update)
DE MIGAPC core protein.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Uehara F., Ohba N., Ozawa M.;
RT "Core Protein of rat Mucinlike Glycoprotein Associated with
RL Photoreceptor Cells.";
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047843; BAB12253.1; -.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 2.
DR SMART; SM00200; SEA; 2.
DR PROSITE; PS00024; SEA; 1.
SQ SEQUENCE 798 AA; 89948 MW; 0CB9DE49182E347C CRC64;

Alignment Scores:
Pred. No.: 1,28e-174 Length: 798
Score: 2504.50 Matches: 516
Percent Similarity: 73.16% Conservative: 92
Best Local Similarity: 62.09% Mismatches: 178
Query Match: 42.63% Indels: 45
DB: 11 Gaps: 11

US-10-007-270-1 (1-3330) x Q9ET62 (1-798)

QY 131 ATGTATTGAAACTAGAGAGCTATTTTGTGTTTGTGTTTCTCCAGTTCAAGCA 190
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 1 MetAsnLeuGluIleLysHisAlaIleLeuValLeuTrpIlePheLeuGlnValGlnGly 20
QY 191 ACCAAAGATATCTCCATTAACTATACCTATCCATTCTGAAACTAAAGACATAGACAAATCCCCCA 250
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
21 IleLysAspThrSerThrLysThrHisSerSerGlyThrLysAsnIleAspLysAlaPro 40
QY 251 AGAATCAAAACAACTGAAAGTACTGAAAAATGTACAAATGTACAAATGTCACTATGAGCAGATA 310
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
41 ArgIleGluThrThrGluSerThrSerAlaValHisLysAlaSerThrMetLysArgLeu 60
QY 311 TTCATTTGGCAAGACATCGAACAACAAAGATCCGCAATTTTCCCAACGGGGGTTAAAGTC 370
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 PheAlaValAlaLysLeuArgAsnLysArgSerAlaLeuPhePro---AlaValAsnIle 79
QY 371 TGCCACAGAAATCCATGAAACAGATTTTAGACAGTCTTCAAGCTTATTATAGATTGAGA 430
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
80 CysProArgGluSerLeuArgGlnIleLeuGluSerLeuGlnIleLysTyrArgLeuArg 99
QY 431 GTGTGTCCAGGAAGCAGATATCGGAAGCATATCGGATCTTCTGATCGCATCCCTGCACACA 490
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
100 ValCysGlnGluIleValTrpGluAlaTyrArgIlePheLeuAspArgValProAspThr 119
QY 491 GGGGAATATCAGGACTGGGTACGATCTCCAGCAGAGAGACCTTCTGCTCTTTTGACATT 550
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
120 GluGluTyrGlnAspTrpValSerLeuCysGlnLysGluThrPheCysLeuPheAspIle 139
QY 551 GGAATAAACTCCCAATCCAGGAGGACCTGATCTTCTCCAGCAGAGATAAAACAG 610
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
140 GlyLysAsnPheSerAsnSerGlnGluHisLeuAspLeuGlnGlnArgIlePheGln 159
QY 611 AGAAGTTTCCCTGCAGAAAGATGAAATATCTGCAGAGAGACATCTGGGAGAGCCGTGT 670
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
160 ArgSerPheSerGlyArgLysAspAspMetSerProIleGluIleLeuGlyValProThr 179
QY 671 GAAACATTCATTTTCAACAGCAATCTACATTTCAAGAGCTTGGGAGATTTCTAAGAA 730
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
180 ThrAlaProValLeu-----
QY 731 AACCTCAGAGAGCAAAATTCAGAGTGTGCCAAGCTCTCACTTGGGCTTTCCCTCTCA 790
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
185 -----ProIle--AspValSerSerMetSerLeuArgProPheProLeuP 199
QY 791 CTCCTGATGACACCTCTCTCAATGAAATCTCGATATACTACATCAACACCAAGATGC 850
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
199 roProAspThrAspLeuLysGluVal-----ThrIleLysAspIleGlnThrP 216
QY 851 CTACACAGAAAGAGAAACAGAAATC-----GCTG 880
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
216 roIleAlaIleArgAlaGluLeuGluSerLysProGluProThrHisValThrGluI 236
QY 881 TGTTCGAGCAGAGAGGTGGAGCTCAGCTCTCTCTGTTAAACACAGAGTTCAAGGCAG 940
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
236 leSerSerGluGluLysValGluPheSerIleSerLeuProAsnHisArgPheLysAla 256
QY 941 ACCTCGTGTACTCCAGTCCCCATTTTACCAGGAGCTAGCAGGAAAGTCCCACTTCA 1000
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
256 luLeuThrAsnSerArgSerProTyrTyrGlnGluLeuValGlyGlnSerGlnLeuGlnL 276
QY 1001 TGCAAAAGATATTAAAGAACTCCAGGATTCAAAATAATCCATGTGTAGATTAGAC 1060
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
276 euGlnLysIlePheLysLysLeuProGlyPheGlyGluIleArgValLeuGlyPheArgp 296
QY 1061 CAAAGAAAGAAAGATCGCTCAAGCTCCACAGAGATGCAACTTACGGCCATCTTTAGA 1120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
296 roLysLysGluGluAspGlySerSerThrGluIleGlnLeuMetAlaIlePheLysA 316
QY 1121 GACACAGTGCAGACCAAAAGCCCTGCAAGTGACCTCTCTCTTTTGTATTTCCAAACAAA 1180
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
316 rGAspHisAlaGluSerLysGlyProGluSerAspLeuLeuSerLeuAspSerAsnLysI 336
QY 1181 TTGAAGTACGAGAACTATCATCTGAGACCATCGAGGAGGACAGCAACAGAAATCTATC 1240
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
336 leGluArgGluArgIleHisGlyAlaIle---GluAspLysGlnProGluAlaTyrL 355

```

|    |      |  |      |
|----|------|--|------|
| QY | 1241 | TCACGCTACAGACCTCAAAGCGCTGATCAGCAAAAGCACTAGAGGAGAAACAATCTTTGG     | 1300 |
| Db | 355  | euThrAlaAaPLeuLysLysLeuIleIleArgLeuLeuAspGlyAspGlnProLeuV        | 375  |
| QY | 1301 | ATGCTGGGACAAITTCAGTTCATCATGATGAATATGCTGGATCACTGCCAGCCCTTTGGCTCGT | 1360 |
| Db | 375  | aIGlyGlyThrValProPheSerAspGluVal-----ThrGluProLeuPheArgProV      | 393  |
| QY | 1361 | ACACCCAATCAGAGCTGCCACATCTTTTGTCTGTATACACAGAGATGCTACTTTGATGC      | 1420 |
| Db | 393  | alThrGlnSerGluLeuProLysProLeuThrAspValThrGluAspValThrLeuSerP     | 413  |
| QY | 1421 | CAGAACTTCTCTGTTGAACCCAGCTTTCAGACAGTGGACGGAGCAGAGCATGCTGTAC       | 1480 |
| Db | 413  | roGluLeuProPheSerGluProArgLeuGlnSerValAspIleIleIleGlyProTyrLeuP  | 433  |
| QY | 1481 | CTGACACTTCTGGTCTCCACTGCTATGGCTCTACTCC---CTGTGAGAAGCTCCAC         | 1537 |
| Db | 433  | roAepSerSerTrpSerArgProValThrAlaSerThrSerGlyValGlyAenLeuProS     | 453  |
| QY | 1538 | CTTTCTTTATGGCATCAAGCATCTTCTCTGACTGATCAAGGACCAAGCATCAATGG         | 1597 |
| Db | 453  | erPhe-----ThrProSerIlePheAlaLeuAspGlnSerSerProLeuMetA            | 471  |
| QY | 1598 | COACTGACACAGACAATGCTAGTACACGGGCTCACATCCCAACAGTGAATATTCGCA        | 1657 |
| Db | 471  | IaThrGlyProThrAlaPheIleProThrLeuThrLeuProIleSerAspIleSerThrV     | 491  |
| QY | 1658 | TCAGCCAACTGGCTCTGGGAATTCACAT---CCACTGCACTTCAGATGACACCGCAT        | 1714 |
| Db | 491  | aArgGlnTrpProLeuGluValSerHisTrpProGlnSerSerSerAspArgGlnLeuS      | 511  |
| QY | 1715 | CAAGTCAGCTGGCGAATATGCTCAGACACTAGATGAATGGAATGCTGCTGACATC          | 1774 |
| Db | 511  | erThrThrSerSerHisaspThrIleArgAspLeuAspGluMetAspValSerAspThrP     | 531  |
| QY | 1775 | CTGCCCACTCAGGTACCGAGCTCAGCGAATATGTTCTGTCCCAATCATTTTCTGG          | 1834 |
| Db | 531  | roAlaLeuSerGluIleAlaGluLeuSerGlyTyrAspSerAlaProAspArgPheLeuG     | 551  |
| QY | 1835 | AGGATACCACTCTCTCAGCTTTACGATATATCAACACTAGTCTTATGACCAATGGCC        | 1894 |
| Db | 551  | luMetThrThrProIleProThrLeuGlnTyrValThr***SerSerGluThrIleAlaA     | 571  |
| QY | 1895 | CCAGAGGCGCAGAGCTGTTAGTGTCTTCAGCTCGCTGTCTCTAACATGCGCTTCTCA        | 1954 |
| Db | 571  | lalyGlyGhiGluLeuValValPhePheSerLeuArgValAlaAenMetProPheSerT      | 591  |
| QY | 1955 | ACGACTGTTTCAACAAGAGCTCTCTGGATACCGAGCTCTGGAGCAACAATTCACACAGC      | 2014 |
| Db | 591  | yzAspLeuPheAenLysSerSerLeuGluTyrGlnAlaLeuGluGlnArgPheThrAspL     | 611  |
| QY | 2015 | TGCTGTCTCCATCTACGATCCATCTTCAGATTATAGCAATCTGAATATCTTAATCT         | 2074 |
| Db | 611  | eueuValProTyrLeuArgSerAenLeuThrGlyPheLysGlnLeuGluIleLeuSerP      | 631  |
| QY | 2075 | TCAGAAACGGAGGTGTATTGTGMAATAGCAAAATGAAGTTTCTTAAGTCTGTGCGGTATA     | 2134 |
| Db | 631  | heArgAsnGlySerValIleValAenSerLysValArgPheAlaLysAlaValProTyrA     | 651  |
| QY | 2135 | ACTCTACCAAGGCTGTGCACGGGGCTTTGGAGGATTTTCTGTCTGACGCCCAACAC         | 2194 |
| Db | 651  | snLeuThrGlnAlaValArgGlyValLeuGluAspLeuArgSerThrAlaAlaGlnGluL     | 671  |
| QY | 2195 | TTCATCTGGAATACAGCTTACTCTCTCAACATTAACACAGCTGATCAACAGAGATCCCT      | 2254 |
| Db | 671  | euaSnLeuGluIleGluSerTyrSerLeuAspIleGluProAlaAspGlnAlaAspProC     | 691  |
| QY | 2255 | GCAAGTTCCTCGCTGGCGGAATTTGCCAATGCTGTAAAGAACGAACGAGCTGAGGAG        | 2314 |
| Db | 691  | ysLysPheLeuAspCysGlyLysPheAaGlnCysIleLysAenGluLeuThrGluGluA      | 711  |

|   |  |   |         |
|---|--|---|---------|
| QY  | 2315   | CGAGTGTGCTGCAAAACGAGATATGACAGCCAGGGAGCGTGCAGCGTCTGGAAACGAC    | 2374    |
| Db  | 711  | laGlucyAeArCyArGInGlyHisGlSerHisGlyThrLeuGluTyGInGluLeuA      | 731     |
| QY  | 2375   | GCCTCTGTGGCCCTGGCACAAGAGNATGCCAGTCTCTCAGGGAAAGGGAGTCCATCGCA   | 2434    |
| Db  | 731  | snLeuCyArProGly---LysThrCysGluAlaSerGInGlyGlnAlaThrProCysA    | 750     |
| QY  | 2435   | GGTTCGCCAGATCACTCTGAAAATCAAGCATACAAACTAGTGTAAAAAGTTCCTCAAAATC | 2494    |
| Db  | 750  | rgProProAspHisSerThrAsnGlnAlaArgGlnProSerValLysLeuGlnArgG     | 770     |
| QY  | 2495   | ACCAAAATACAGAGTAATCAGTAAAGAATTTCTGAATTACTGACGCTAGAAATATGAAG   | 2554    |
| Db  | 770  | InGln---AsnLysValLysLysArgAsnSerGluLeuSerAlaThrAspPheGluG     | 789     |
| QY  | 2555   | AATTTAAACCATCAAGATTGGGAAGCAAT                                 | 2583    |
| Db  | 789  | luLeuAspAspGlnAspTrpGluGlyAsn                                 | 798     |
| RESULT 5                                  |  |   |         |
| QY  | Q8RIW8   | PRELIMINARY;  | 798 AA. |
| AC  | Q8RIW8;  |   |         |
| DT  | 01-JUN-2002 (TrEMBLrel. 21, Created)                               |   |         |
| DT  | 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)                  |   |         |
| DT  | 01-JUN-2003 (TrEMBLrel. 21, Last annotation update)                |   |         |
| DE  | Interphotoreceptor matrix proteoglycan 1.                          |   |         |
| GN  | IMPG1.   |   |         |
| OS  | Mus musculus (Mouse).  |   |         |
| OC  | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;            |   |         |
| OC  | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. |   |         |
| OX  | NCBI_TaxID=10090;  |   |         |
| RN  | [1]  |   |         |
| RP  | SEQUENCE FROM N.A.   |   |         |
| RC  | TISSUE=Eye;  |   |         |
| RA  | Strausberg R.;   |   |         |
| RL  | Submitted (FEb-2002) to the EMBL/GenBank/DBJ databases.            |   |         |
| DR  | EMBL, RC022970, AH22970.1; -.                                      |   |         |
| DR  | MCD; MG1:1926876; Impg1.   |   |         |
| DR  | GO; GO:0004872; Fireceptor activity; IEA.                          |   |         |
| DR  | InterPro; IPR000082; SEA_domain.                                   |   |         |
| DR  | Pfam; PF01390; SEA; 2.   |   |         |
| DR  | SMART; SM00200; SEA; 2.  |   |         |
| KW  | Receptor.  |   |         |
| SQ  | SEQUENCE 798 AA; 89474 MW; E875FB5587389D4C CRC64;                 |   |         |
| Alignment Scores:                         |  |   |         |
| Pred. No.:                                | 1.8e-174   | Length:   | 798     |
| Score:                                    | 2502.50  | Matches:  | 527     |
| Percent Similarity:                       | 73.13%   | Conservative:   | 80      |
| Best Local Similarity:                    | 63.49%   | Mismatches:   | 180     |
| Query Match:                              | 42.60%   | Indels:   | 43      |
| DB:                                       | 11   | Gaps:   | 10      |
| US-10-007-270-1 (1-3330) x Q8RIW8 (1-798) |  |   |         |
| QY  | 131  | ATCATTTTGGAACTAGACAGACTATTTTGTGTTTGGATTTTCTCCAGTTCAGGA        | 190     |
| Db  | 1  | MetAsnLeuGlnIleLysHisAlaIlePheValLeuGlyIlePheLeuGlnValGIngly  | 20      |
| QY  | 191  | ACCAAGATATCTCCATTAACATATACCATTCTGAACTAAGACATAGACATCCCCCA      | 250     |
| Db  | 21   | IleLysAspThrSerIleLysIlePheSerSerGluIleLysAsnIleAspLysThrPro  | 40      |
| QY  | 251  | AGAAATGAACCACTGAAAGTACTGAAAAATGTCAAAAATGTCAACTATGACGAGATA     | 310     |
| Db  | 41   | ArgIleGluThrIleGluSerThrSerThrValHisLysValSerThrMetLysArgIle  | 60      |
| QY  | 311  | TTTCGATTTGGCAAGCATCGACAAAGATCCGCAATTTTCCCAACGGGGTTAAAGTC      | 370     |
| Db  | 61   | PheAspLeuProLysLeuArgThrLysArgSerAlaLeuPhePro---AlaAlaAsnIle  | 79      |



Db 770 InGlnAenLysValValLysLysArgAsnSerLysLeuSerAlaIleGlyPheGluGluP 790  
 QY 2558 TTAACCATCAAGATTGGAGGAAT 2583  
 Db 790 heGluaspGlnAspTrpGluGlyAsn 798  
 RESULT 6  
 Q9ES62 PRELIMINARY; PRT; 742 AA.  
 AC Q9ES62;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Sialoprotein associated with cones and rods SPACR.  
 GN IMPG1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Lee J.W., Chen Q., Rayhorn M.E., Shadrach K.G., Crabb J.W.,  
 RA Rodriguez I.R., Hollyfield J.G.;  
 RT "SPACR in the interphotoreceptor matrix of the mouse retina: molecular  
 RT and biochemical characterization";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF229929; AAC32162.1; --  
 DR MGD; MGI:1926876; Impgl.  
 DR InterPro; IPR000082; SEA\_domain.  
 DR Pfam; PF01390; SEA; 2.  
 DR SMART; SM00200; SEA; 2.  
 SQ SEQUENCE 742 AA; 83177 MW; 58CAOCFE0FOA9B97 CRC64;

Alignment Scores:

Pred. No.: 7,66e-163 Length: 742  
 Score: 2343.50 Matches: 494  
 Percent Similarity: 73.13% Conservative: 72  
 Best Local Similarity: 63.82% Mismatches: 165  
 Query Match: 39.89% Indels: 43  
 DB: 11 Gaps: 10

US-10-007-270-1 (1-3330) x Q9ES62 (1-742)

QY 299 ATGAGCAATATTCGATTGGCAAGCATCGAACAAGATCCGATTTTCGCAAG 358  
 Db 1 MetLysArgIlePheAspLeuProLysLeuArgThrLysArgSerAlaLeuPhePro--- 19  
 QY 359 GGGTTAAAGTCTGTCCACAGATCCATCAACAGATTTTACAGACTTCAAGCTTAT 418  
 Db 20 AlaAlaAsnIleCysProGlnGluSerLeuArgGlnIleLeuAlaSerLeuGlnGluTyr 39  
 QY 419 TATAGATTGAGAGTGTGTCAGGAGCAGTATGGAGCATATCGGATCTTTCTGATGCG 478  
 Db 40 TyrArgLeuArgValCysGlnGluValValTrpGluAlaTyrArgIlePheLeuAspArg 59  
 QY 479 ATCCCTGACACAGGGGATATCAGGACTGGGTGAGCATCTGCCAGCAGACACCTTCG 538  
 Db 60 IleProAspThrGluGluTyrGlnAspTrpValSerLeuCysGlnLysGluThrPheCys 79  
 QY 539 CTCTTTGACATGGAAAAAATCTTCAGCAATTCAGGAGCAGCTGATCTTCTCCAGCAG 598  
 Db 80 LeuPheAspIleGlyLysAsnPheSerAsnSerGlnGlnIleLeuAspLeuLeuGlnGln 99  
 QY 599 AGAATAAAACAGAGAGTTCCCTGACAGAAAGATGAATATCTGCGAGAGAGACATTG 658  
 Db 100 ArgIleGlnArgSerPheProGlyArgLysAspGluThrAlaSerMetGluThrLeu 119  
 QY 659 GGAGAGCCTGGTGAACCAATTTGATTCACAGCAATCTACATTTCAAGACTTGGGCA 718  
 Db 120 GluAlaProThrGluAlaProValPro--Thr----- 130  
 QY 719 GTATTTCAAGAAAAACCTCTCAGAGAGCAAAATTCAGATGTTGCCAATCTCTACCTTGGGC 778  
 Db 1 MetLysArgIlePheAspLeuProLysLeuArgThrLysArgSerAlaLeuPhePro--- 19

Db 131 -----AspValSerArgMetSerLeuGlyP 139  
 QY 779 CTTTCCCTCTCACTCTGATGACACCTCTCTCAATGAAATTCGATATACATCAACG 838  
 Db 139 roPheProLeuProSerAspAspThrAspLeuLysGluIleLeuSerValThrLeuLysA 159  
 QY 839 ACACCAAGATGCTTACACAGAAAGAGAAACA-----CAATTCCTG 880  
 Db 159 spIleGlnLysProThrThrGluSerIleThrGluProIleHisValSerGluPheSer- 178  
 QY 881 TGTGGAGGAGCAGAGAGGTGGAGCTCAGCGTCTCTCTGGTAAACACAGAAAGTTCAAGCAG 940  
 Db 179 -----SerGluLysValGluPheSerIleSerLeuProAsnHisArgPheLysAlaG 197  
 QY 941 AGTCGCTGACTCCCTCCCTCCCTATATTACAGAGAGTACGAGAAAGTCCCACTTCA 1000  
 Db 197 lueUthrAsnSerGlySerProTyrGlnGluLeuValGlnGlnSerGlnLeuGlnL 217  
 QY 1001 TGCAAAAGATATTAAAGAAATTCAGGATTCAAAATAATCCATGTGTGATGATTTAGAC 1060  
 Db 217 euGlnLysIlePheLysLeuProGlyPheGlyGluIleArgValLeuGlyPheArgP 237  
 QY 1061 CAAGAAGAAAAGATGGCTCAAGCTCCACAGAGATGCAACTTACGGCCATCTTTAAGA 1120  
 Db 237 roLysLysGluLysAspGlySerSerSerThrGluIleGlnLeuMetAlaIlePheLysA 257  
 QY 1121 GACACAGTGCAGAGAGAAAGCCCTGCAAGTCACTGCTCTCTTTGATTTCCAAACAAA 1180  
 Db 257 rgAspHisAlaGluAlaLysSerProAspSerHisLeuSerLeuAspSerAsnLysI 277  
 QY 1181 TTGAAAGTGAAGAGTCTATCATGCAACCTGAGAGAGCAGACAGCAACCAAGATCTATC 1240  
 Db 277 leGluSerGluArgIleHisGlyValIle---GluAspLysGlnProGluThrTyrL 296  
 QY 1241 TCACAGCTACAGACTCAAAAGCTGATCAGCAAGACACTAGAGAGAGAAACAATCTTTGG 1300  
 Db 296 eutrAlaThrAspLeuLysLysLeuIleGlnLeuLeuAspGlyAspLeuSerLeuV 316  
 QY 1301 ATGTGGGACAAATTCAGTTCACTGATGAATTCCTGATCACTGCGACCTTTGCTCTG 1360  
 Db 316 alGluGlyLysIleProPheGlyAspGluValThrGlyThrLeu-----PheArgProV 334  
 QY 1361 ACACCAATCAAGAGTGCCTCCATCTTTCTGCTTATACAGAGAGTCTACTTTGATGTC 1420  
 Db 334 alThrGluProAspLeuProLysProLeuAlaAspValThrGluAspAlaThrLeuSerP 354  
 QY 1421 CAGACTTCTCTCTGTTGAAACCCAGCTTGAGACAGTGGACGAGCAGAGCATGCTCTAC 1480  
 Db 354 roGluLeuProPheValGluProArgLeuGluAlaValAspArgGluGlySerGluLeuP 374  
 QY 1481 CT-----GACACTTCTTGGTCTCCACCTGCTATGGCTCTACTCTCTCTCT 1525  
 Db 374 roGlyMetSerSerLysAspSerTrpSerProValSerAlaSerIleSerArgS 394  
 QY 1526 CAGAGTCCACTTCTTTTATGCAAGCATCTTCTCTGACTGATCAAGGCCACCA 1585  
 Db 394 exGluAsnLeuProSerPhe---ThrProSerIlePheSerLeuAspAlaGlnSerProp 413  
 QY 1586 CAGATACAATGGCCCTCAGCAGACATCTAGTACAGGGGCTACCATCCCAACCAAGT 1645  
 Db 413 roProLeuMetThrThrGlyProThrAlaLeuIleProLysProThrLeuProThrIleA 433  
 QY 1646 ATTATTCGCAATCAGCCAACTGGCTCTGGGAATTCACATCCACCTGCTCAGATG 1705  
 Db 433 epTyrSerThrIleArgGlnLeuProGluSerHisTrpProAlaSerSerSerA 453  
 QY 1706 ACAGCCCATCAAGTCCAGTGGCGAAGATATGGTCAGACACCTAGATGAATGATCTGT 1765  
 Db 453 spArgGluLeuIleThrSerSerHisAspThrIleArgAspLeuAspGlyMetAspValS 473  
 QY 1766 CTGACACTCTCCCTCTGAGTACCGAGCTCAGCGAATATGTTCTCTCCAGATC 1825  
 Db 473 erAspThrProAlaLeuSerGluLeuSerGlyTyrAspSerAlaSerGlyC 493



QY 1826 ATTTCTTGAGGATACCACTCTCTGCTCAGCTTACAGTATATACCACTAGTTCTATGA 1885  
 Db 493 lnPheLeuGluMetThrThrProIleProThrValAlaGpPheIleThrThrSerSerGluT 513  
 QY 1886 CCATTCCCGCCCAAGCGGAGAGCTGGTAGTGTCTTCTCAGTCTGGCTGTGCTTAACATGG 1945  
 Db 513 hrIleAlaThrlyGlyGlnGluLeuValPhePheSerLeuValAlaAsnMetP 533  
 QY 1946 CTTTCTCCACGACCTGTTCAACAGAGCTCTCTGGAGTACCGAGCTCTGAGCAACAT 2005  
 Db 533 roPheSerTyrAspLeuPheAsnLysSerSerLeuGluTyrGlnAlaLeuGlnGlnArgP 553  
 QY 2006 TCACACAGCTGCTGGTTCCTATCTACGATCCAACTTACAGATTAAAGCACTTGAAA 2065  
 Db 553 heThrAspLeuLeuValProTyrLeuArgSerAsnLeuThrGlyPheLysGlnLeuGluI 573  
 QY 2066 TACTTAACCTTCAGAAACGGAGGTGATTGTGAATAGCAAAATGAAGTTGCTAGTCTG 2125  
 Db 573 leLeuSerPheArgAsnGlySerValIleValAsnSerLysValArgPheAlaLysAlaV 593  
 QY 2126 TGGCGTTAACTCACCAAGGCTGTGCACGGGTCTTGGAGGATTTCTGCTGCTGAG 2185  
 Db 593 alProTyrAsnLeuThrGlnAlaValAlaArgGlyValLeuGluAspLeuArgSerThrAla 613  
 QY 2186 CCCAACAACTCCATCGGAATAGACAGCTACTCTCTCAACATTGAACAGCTGATCAAG 2245  
 Db 613 laGlnGlyLeuAsnLeuGluIleGluSerTyrSerLeuAspIleGluProAlaAspGln 633  
 QY 2246 CAGATCCCTGCAAGTTCTCTGGCTCGCGGCAATTCCTCCATGTGTAAAGAACGACGGA 2305  
 Db 633 laAspProCysLysLeuLeuAspCysGlyLysPheAlaGlnCysValLysAsnGluTPT 653  
 QY 2306 CTGAGGAGCGAGTGTCTGCTCAACACGAGTATGACAGCGGAGCGCTGGACGCTC 2365  
 Db 653 hrGluGluAlaGluCysArgCysArgGlnGlyHisGluSerHisGlyThrLeuAspTyrG 673  
 QY 2366 TGGAACACGAGGCTCTGTGGCTGTGCACAAAGCAAGCAAGTCCCTCCAGGAAGGAG 2425  
 Db 673 lnThrLeuAsnLeuCysProGly---LysThrCysValAlaGlyArgGluGlnAlaT 692  
 QY 2426 CTCCATGAGGTTGCCAGATCACTCTGAAATCAAGCATACAAACTAGTGTAAAAAGT 2485  
 Db 692 hrProCysArgProProAspHisSerThrAsnGlnAlaGlnGluProGlyValLysLysL 712  
 QY 2486 TCCAAATCAACAAATACAGAGTATCAGTAAAGAAATCTGAAATCTGACCGTAG 2545  
 Db 712 eu-----ArgGlnGlnAsnLysValLysLysArgAsnSerLysLeuSerAlaIleG 730  
 QY 2546 AATATGAAGAAATTAACCATCAAGATTGGGAAGGAAAT 2583  
 Db 730 lyPheGluGluPheGluAspGlnAspTrpGluGlyAsn 742

RESULT 7  
 Q8JIR8  
 ID Q8JIR8 PRELIMINARY; PRT; 928 AA.  
 AC Q8JIR8;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE SPACR.  
 GN CSFACR.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archoosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Retina;  
 RX MEDLINE=22095579; PubMed=11991949;  
 RA Zako M., Iwaki M., Yoneda M., Miyaiishi O., Zhao J., Suzuki Y.,  
 RA Takeuchi M., Miyake G., Ikagawa H., Kimata K.;

RT "Molecular Cloning and Characterization of Chick Sialoprotein  
 Associated with Cones and Rods, a Developmentally Regulated  
 Glycoprotein of Interphotoreceptor Matrix.";  
 U. Biol. Chem. 277:25592-25600(2002).

DR EMBL; AB070714; BAC00947.1; -  
 DR InterPro; IPR006209; EGF like.  
 DR InterPro; IPR000082; SEA\_domain.  
 DR Pfam; PF01390; SEA; 2.  
 DR SMART; SM00200; SEA; 2.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS00024; SEA; 1.  
 SQ SEQUENCE 928 AA; 102684 MW; 4DBB800528C56CA2 CRC64;

## Alignment Scores:

Pred. No.: 9,53e-124 Length: 928  
 Score: 1810.00 Matches: 414  
 Percent Similarity: 55.68% Conservatives: 120  
 Best Local Similarity: 43.17% Mismatches: 199  
 Query Match: 30.81% Indels: 226  
 DB: 13 Gaps: 18

US-10-007-270-1 (1-3330) x Q8JIR8 (1-928)

QY 131 ATGTATTGGAACACTAGAGAGCTATTTTGTGTTTGGATTTTCTCCAGTTCAAGA 190  
 Db 1 MetHisLeuLysThrGlyLeuIlePheLeuAlaIleCysLeuAlaLeuGlnValGlnGly 20  
 QY 191 ACCAAAGATATCTCCATTACATATACCATCTGCAACTAAAGACATAGACATATCCCA 250  
 Db 21 SerArgGluIleProSerLysThrAsnHisGlyGluAlaLysGlnLeuAlaAspAlaSer 40  
 QY 251 AGAAATGAACAACTGAAAGTACTGAAAAAATGTACAAAATGTCACTATGACGACCAATA 310  
 Db 41 GlySerAspLysThrGluArgThrThrLysArgSerArgValSerThrIleArgGly 60  
 QY 311 TTGATTGGCAAGCATCGAACAAGATTCGGCATTTTCCCAACGGGGTTAAAGTC 370  
 Db 61 PheAspMetAlaLysHisArgThrLysArgSerProPhePheSerThrGlyValLysIle 80  
 QY 371 TGTCACAGGAATCCATGAAACAGATTTTACAGCTTCTCAGCTTATATATAGATTGGA 430  
 Db 81 CysProGlnGlnSerValLysGlnIleLeuAlaSerHisGlnAlaTyrTyrArgLeuArg 100  
 QY 431 GTGTCTCAGGAAGCAGTATGGAGCATATCGGATCTTTCTGGATCGCATCCCTGACACA 490  
 Db 101 ValCysGlnGluAlaValTrpGluAlaPheArgIlePheLeuAspArgIleProAspThr 120  
 QY 491 GGGGAATATCAGGACTGGGTTCAGCATCTGCACAGGAGACCTTCTCCCTTTGACATT 550  
 Db 121 SerGluTyrGlnAsnTrpValThrAlaCysGlnArgGluThrPheCysIlePheAspIle 140  
 QY 551 GGAAAAAACTTCAGCAATTCACAGAGCAGCTGGATCTTCTCCAGCAGAGATAAAACAG 610  
 Db 141 GlyLysAsnPheSerAsnSerGlnGluHisGluGluIleGlnArgArgValLysHis 160  
 QY 611 AGAAGTTCCCTGCAGAAAGATGAATATCTGCAGAGAAGACATTTGGAGAGCGCTCGT 670  
 Db 161 ArgThrPheGlnGluArgLysAspGluIleSerThrAspLysThrGlyGly----- 177  
 QY 671 GAAACCATGTCTATTCACAGCAATCTACATTTCAAGACTTGGCGAGTATTTCTAAGAA 730  
 Db 177 ----- 177  
 QY 731 AACCTCAGAAGACAAATTCAGATGTTGCCAAGCTCTCACTTGGGCT----- 780  
 Db 178 -----Lys--LysLeuGluAspIleProSerValSerThrGlyProProSerAlaAs 194  
 QY 781 -----TTCCCTCTCTCTCTGATGACACCTCTCTCAATGAATTTCTCGATAATA 829  
 Db 194 erLeuSerThrTyrThrLeuValProAsnGlyThrLeuLeuAsnGluIle----- 210  
 QY 830 CACTCAACGACACCAAGATGCTACACACGAAAGAAACAGAAATTCGCT-----GTGT 883  
 Db 830 -----GTGT-----



Db 211 --ValAsnGluThrLysThrProValLysGluLeuGlyThrAsnThrValProGluLeuP 230  
 QY 884 TGGAGGACGAGGGTGGAGCTCAGCTCTCTGTGTAAACAGAGGTTCAAGCGCAGAGC 943  
 Db 230 roAlaGluGlnMetValGluPheSerValThrLeuThrAspGlnGlyThrAlaGluL 250  
 QY 944 TCCTGACTCCCGCTCCCATATTACAGGAGCTACAGGAAGTCCCACTTCAGATGC 1003  
 Db 250 euSerAspProAsnSerProGlnThrArgGlnLeuAlaAlaPheGlnLeuGlnMetG 270  
 QY 1004 AAAGATATTAAAGAACTTCCAGGATTCAAAAAATCCATGTGTAGGATTTAGACAA 1063  
 Db 270 InLysIlePheGluLysLeuProGlyPheLysGlnIleHisValLeuGlyPheLysGlnL 290  
 QY 1064 AGAAGAAAGAGTGGCTCAAGCTCCAGAGATGCAACTTACGGCCATCTTTAAGAGAC 1123  
 Db 290 YbLysGluLysAspGlySerSerThrIleAlaArgTyrMetValAsnPheGluArg 310  
 QY 1124 ACAGTGCAGAAAGAAAAGCCCTGCAAGTGACCTCTGTCITTTGATTCACAAATG 1183  
 Db 310 spGlySerGluLysSerThrAlaAspAspIleSerThrIleGlySerAsnLysValG 330  
 QY 1184 AAAGTGAAGAGTCTATCATGGAACCATGGAGGACAAAGCAACCAAGAAATCTATCTCA 1243  
 Db 330 LuAsnGluLysValProLeuSerAlaLysGluLysArgGluLysSerAlaThrLysLeuT 350  
 QY 1244 CAGCTACAGACCTCAAAAGCTGATCAGCAACACTAGAGAGAACAACTTTG----- 1299  
 Db 350 hrValThrAspLeuGlnGlnLeuValAlaThrAlaLeuHisGluAspArgSerLeuProV 370  
 QY 1300 --GATGGGGACAATTCACTTCACTCATGAA----- 1329  
 Db 370 alAspLeuGlyThrLeuArgPheThrAspGluProIleLysProSerSerAspPheAspA 390  
 QY 1330 -----ATTGCTGA-----TCACTGCCAGCCTTTGGCTGAC-----ACCC 1366  
 Db 390 enAspIleGlnGlyMetValThrIleProLeuAlaGlyProAspLeuAspAspThrIleS 410  
 QY 1367 AATCAGAGCTGCCCATCTTT----- 1389  
 Db 410 erAlaGluLeuProLeuValTyrProSerProIleThrValAspGlnThrArgAspIleP 430  
 QY 1389 ----- 1389  
 Db 430 heValAspGluPheThrThrGlyIleThrAspLeuSerArgGluIleGlyGlyProGluA 450  
 QY 1389 ----- 1389  
 Db 450 spPheAspSerAsnPheIleThrSerGluProAlaPheProThrLysProSerArgGluP 470  
 QY 1389 ----- 1389  
 Db 470 roProHisAspArgSerProAspThrGluAspIleThrAspTyrGlnArgPheThrV 490  
 QY 1380 -----GCTGTATACAGAGATGCTACTTTGAGTCCAGAA-----CTTC 1429  
 Db 490 alProPheSerAlaLeuValSerThrAspSerProAlaLysProGluAspSerTyrLeuP 510  
 QY 1430 CTCTGTTGAACCCGCTTGAGACAGTGGAGGAGCAGACATGGTCTACCTGACACTT 1489  
 Db 510 roProProAlaAspGluSerAspSerAsnAspLeuIleThrAspGluSerProThrGluG 530  
 QY 1490 CTGGTCTCAGCTCTGAGCTCTACCTCCCTGTCAGAGAGTCCACTTCTTCTTATGG 1549  
 Db 530 InValIleThrProAlaValTyrThrThrGlySerPheThrLeuProThrPheLeuGlnA 550  
 QY 1550 CATCAAGC-----A 1558  
 Db 550 laThrAspLysAspThrGluAlaGluMetLysLysGluLeuValGlyValThrGluProL 570  
 QY 1559 TCTTCTCTGACTGATCA----- 1578  
 Db 570 eupHelysGluAlaAspArgSerLeuSerGlyGlnAlaValLysMetMetAspGluL 590

QY 1578 ----- 1578  
 Db 590 euGluSerSerGlyAspAspIleLeuValThrThrSerThrTyrLysThrLeuProPheL 610  
 QY 1579 -----GGCACACACATCAATGGCCACT----- 1602  
 Db 610 euIleGlySerSerPheAlaThrGlnProGluValThrPheAlaAlaLeuP 630  
 QY 1603 -----GACCAGACATCTAGTACAGGGCTCACCATCCACCAGTGAATTTCTGCAA 1657  
 Db 630 roProAspGlnThrLeuLeu---ProThrValThrSerProPheTyrSerHisSerValV 649  
 QY 1658 TCAGCCAACTGGCTCTCGGAATTTCCATCCACCT----- 1692  
 Db 649 al-----IleAspGlnSerProGluValProAspThrLeuMetProA 663  
 QY 1693 -----GCATCTTCAGATGACAGCCGATCAAGTCAGGTGGCGAAGATATGTCAGACAC 1747  
 Db 663 laAlaAlaSerAlaLeuProAspArgAlaSerThrGlyValGlnAspIleAlaGluL 683  
 QY 1748 TAGATCAATGATCTGTCTGACATCTCTGCCCATCTGAGTACCGAGCTCAGGAAT 1807  
 Db 683 euAspGlyAlaGlyValLysSerThrAlaValLeuAspGluAlaGluHisGlySerGlyT 703  
 QY 1808 ATGTTTCTGCTCCAGATCATTTCTTGGAGGATACCATCTCTCTCAGCTTTTACAGTATA 1867  
 Db 703 YrIleSerValGlnThrThrGluProAlaGluValThrGlnAlaProThrLeuLysYrV 723  
 QY 1868 TCACCATGATTTCTATGACCATTTGCCCGGCGAGAGCTGTAGTGTCTTCTTCAGTC 1927  
 Db 723 alThrSerSerMetThrThrAlaAlaLysGlyLysGluLeuValValPhePheSerL 743  
 QY 1928 TGGGTTGCTTAACATGGCTTCTCCACAGCACTTTTCAACAGAGCTCTCTGAGATACC 1987  
 Db 743 euArgValThrAsnMetHisPheSerAspLeuPheAsnArgSerSerGlnGluTyrL 763  
 QY 1988 GAGCTCTGGAGCAACAATTCACACAGCTGTGTTCCATCTCATCTAGCTCAATCTTACAG 2047  
 Db 763 ysAlaLeuGluGlnGlnPheMetGlnLeuLeuLeuProTyrLeuGlnSerAsnLeuThrG 783  
 QY 2048 GATTTAGCAACTTGAATTAATTAATTCAGAAACGGAGTGTGATTTGTAATAGCAAAA 2107  
 Db 783 lypHelysGlnLeuGluIleLeuAsnPheArgAsnGlySerValIleValAsnSerLysM 803  
 QY 2108 TCAAGTTTGTAACTGTGTGCGGTATTAACCTCACCAAGCTGTGCACGGGTCTTGGAGG 2167  
 Db 803 eLysPheAlaArgThrValProTyrAsnIleThrGluAlaValHisCysValLeuGluA 823  
 QY 2168 ATTTTCTGTTCTGTCAGCCCAACAACTCCATCTGGAATAGACAGCTACTCTCTCAACA 2227  
 Db 823 epPheCysAspAlaAlaAlaGlnHisLeuAsnLeuGluIleAspSerTyrSerLeuAspI 843  
 QY 2228 TTGACACAGCTGATCAGCAGATCCCTGCAAGTCTCTGGCTCGCGCAATTTGCCAAT 2287  
 Db 843 leGluProAlaAspGlnAlaAspProCysLysPheMetAlaCysAspGluPheSerLysC 863  
 QY 2288 GTGTAAGAAACGACCGAGCTCAGGACGGAGTGTGCTGCCAACACAGCATATGACGCC 2347  
 Db 863 ysIleMetAsnGluTrpThrLysGluAlaAspCysLeuCysLysProGlyTyrAlaSerG 883  
 QY 2348 AGGGG-----AGCCTGGACGGTCTGGAACACGAGCTCTGTGGCCCTGGCA 2392  
 Db 883 InAspGlyLeuProCysArgSerLeuCysGluMetGluProHisLeuCysAspAsnGlyG 903  
 QY 2393 CAAGAGAAATCGAGTCTCTCCAGGAAAGGAGTCCATCATCAGTTGCCAGAT 2445  
 Db 903 lyllys---CysGluLeuValProGlyArgGlyAlaValCysArgSerProAsp 919  
 RESULT 8  
 ID Q9BWZ1  
 AC Q9BWZ1; PRELIMINARY; PRT; 288 AA.

| Db   | 188  | -----AspValAlaAsnValSerLeuGlyProPheProLeuT                         | 200           |  |  |
|--|--|--|---------------|--|--|
| QY   | 791  | CTCCCTGATGACACCTTCCTCAATGAATTCCTGATAATACATCAACGACACCAAGATGC        | 850           |  |  |
| Db   | 200  | hrProAspThrLeuLeuAsnGluIleLeuAspAsnThrLeuAsnAspThrIlysMetP         | 220           |  |  |
| QY   | 851  | CTACACACAGAAAGAGAAACAAATTCGCTGCTGTCGAGGAGCAGAGGGTGGAGCTCAGCG       | 910           |  |  |
| Db   | 220  | roThrThrGluArgGluThrGluPheAlaValIleuGluGluGlnArgValGluLeuSerV      | 240           |  |  |
| QY   | 911  | TCTCTCTGTTAAACACAGAGTTCAAGCAGAGCTCGCTGACTCCAGTCCCCCATATTACC        | 970           |  |  |
| Db   | 240  | alSerLeuValAsnGlnIlysPheIysAlaGluLeuAlaAspSerGlnSerProIlyrTrg      | 260           |  |  |
| QY   | 971  | AGGAGCTACAGAGAAAGTCCCACTTCAGATCCAAAGATATTATTAAGAACTTCCAGGAT        | 1030          |  |  |
| Db   | 260  | InGluLeuAlaGlySerGlnLeuGlnMetGlnIlylePheIysIysLeuProGlyP           | 280           |  |  |
| QY   | 1031   | TCAAAAAATCCATGTTAGGATTTT   | 1056          |  |  |
| Db   | 280  | heLysIlylleHisValLeuGlyPhe   | 288           |  |  |
| RESULT 9                                   |  |  |               |  |  |
| Q9UKK5                                     |  |  |               |  |  |
| ID   | Q9UKK5   | PRELIMINARY;   | PRT; 1241 AA. |  |  |
| AC   | Q9UKK5;  |  |               |  |  |
| DT   | 01-MAY-2000  | (T-EMBLrel. 13, Created)   |               |  |  |
| DT   | 01-MAY-2000  | (T-EMBLrel. 13, Last sequence update)                              |               |  |  |
| DT   | 01-OCT-2003  | (T-EMBLrel. 25, Last annotation update)                            |               |  |  |
| DE   | Sparcan.   |  |               |  |  |
| OS   | Homo sapiens (Human).  |  |               |  |  |
| OC   | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;              |  |               |  |  |
| OC   | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.           |  |               |  |  |
| OX   | NCBI_TaxID=9606;   |  |               |  |  |
| RP   | [1]_TaxID=9606;  |  |               |  |  |
| RN   | SEQUENCE FROM N.A.   |  |               |  |  |
| RE   | MEDLINE=201671166; PubMed=10702356;                                  |  |               |  |  |
| EX   | Acharya S., Foletta V.C., Lee J.W., Rayborn M.E., Rodriguez I.R.,    |  |               |  |  |
| RA   | Young W.S. III, Hollyfield J.G.;                                     |  |               |  |  |
| RT   | "SPACRAN, a novel human interphotoreceptor matrix hyaluronan-binding |  |               |  |  |
| RT   | proteoglycan synthesized by photoreceptors and pinealocytes.";       |  |               |  |  |
| RL   | J. Biol. Chem. 275:6945-6955(2000).                                  |  |               |  |  |
| DR   | EMBL; AF157624; AAF13154.1; -  |  |               |  |  |
| DR   | GO; GO:0005578; C:extracellular matrix; TAS.                         |  |               |  |  |
| DR   | GO; GO:0005201; F:extracellular matrix structural constituent; TAS.  |  |               |  |  |
| DR   | GO; GO:0005540; F:hyaluronic acid binding; TAS.                      |  |               |  |  |
| DR   | GO; GO:0007601; P:vision; TAS.                                       |  |               |  |  |
| DR   | InterPro; IPRO06209; EGF like.                                       |  |               |  |  |
| DR   | InterPro; IPRO00082; SEA_domain.                                     |  |               |  |  |
| DR   | Pfam; PF01330; SEA; 2.   |  |               |  |  |
| DR   | SMART; SM00200; SEA; 2.  |  |               |  |  |
| DR   | PROSITE; PS01186; EGF_2; 1.  |  |               |  |  |
| DR   | PROSITE; PSS0024; SEF; 1.  |  |               |  |  |
| SQ   | SEQUENCE 1241 AA; 138606 MW; 1F3AE63DB39F8858 CRC64;                 |  |               |  |  |
| Alignment Scores:                          |  |  |               |  |  |
| Pred. No.:                                 | 9,24e-59   | Length:  | 1241          |  |  |
| Score:                                     | 923.00   | Matches:   | 316           |  |  |
| Percent Similarity:                        | 39.15%   | Conservative:  | 162           |  |  |
| Best Local Similarity:                     | 25.88%   | Mismatches:  | 312           |  |  |
| Query Match:                               | 15.71%   | Indels:  | 431           |  |  |
| DB:  | 4  | Gaps:  | 38            |  |  |
| US-10-007-270-1 (1-3330) x Q9UKK5 (1-1241) |  |  |               |  |  |
| QY   | 170  | ATTTTCTCCCAATCAAGAAC--AAAGATATCTCCATTAACATATACCATTCGAA             | 226           |  |  |
| Db   | 16   | IlePheValLeuIleGluGlyAspPheProSerLeuThrAlaGlnThrIlyrLeuSer---      | 34            |  |  |
| QY   | 227  | ACTAAGACATAGACAAATCCC-----CCAAAGAAATGAACA                          | 262           |  |  |
| Db   | 35   | IleGluIleGlnIleGlnIleProIlyrLeuSerAlaValSerPheLeuLeuProGluGluSerTh | 54            |  |  |

|      |    |  |
|------|----|--|
| 1288 | QY | -----GAACAATCTTTCGATGTG- 1305                                      |
| 377  | Db | roAspSerLeuGlnLeuIleAsnValArgGlyValleuArgHisGlnThrGluAspLeuV 397   |
| 1306 | QY | -----GGACAATTCAGTCTACT- 1323                                       |
| 397  | Db | alTrpAsnThrGlnSerSerSerLeuGlnAlaThrProSerSerIleLeuAspAsnThrP 417   |
| 1324 | QY | -----GATGAATTCGTGGATCACTCCA- 1347                                  |
| 417  | Db | heGlnAlaalaTrpProSerAlaAspGluSerIleThrSerSerIleProLeuAspP 437      |
| 1348 | QY | -----GCCTTTGGTCCTGCACACCAATCAGACTGCC- 1380                         |
| 437  | Db | heSerSerGlyProProSerAlaThrGlyArgGluLeuTrpSerGluSerProLeuGlyA 457   |
| 1380 | QY | ----- 1380   |
| 457  | Db | spLeuValSerThrHisLysLeuAlaPheProSerIysMetGlyLeuSerSerSerProG 477   |
| 1381 | QY | -----ACAT 1384   |
| 477  | Db | luValleuGluValSerSerLeuThrLeuHisSerValThrProAlaValleuGlnThr 497    |
| 1385 | QY | CTTTTGCTGCTATTAACAGAGAGTCTACTTTGAGTCCAGACTTCCTCGTTGAACCCC 1444     |
| 497  | Db | lyLeuProValAlaIleSerGluGluArgThrSerGlySerHisLeu-----ValGluAspG 515 |
| 1445 | QY | AGCTTTGACAGACGAGCGGACGACAG- 1489                                   |
| 515  | Db | lyLeuAlaAsnValGluGluSerGluAspPheLeuSerIleLeuAspSerLeuProSerS 535   |
| 1490 | QY | CTTGCTCTCCACCTGCTATGCGCTCTACC- 1522                                |
| 535  | Db | erPheThrGlnProValProLysGluThrIleProSerMetGluAspSerAspValSerL 555   |
| 1523 | QY | TGTCAGAGACTCA- 1564  |
| 555  | Db | euThrSerSerProTyLeuThrSerSerIleProPheGlyLeuAspSerLeuThrSerL 575    |
| 1565 | QY | CTCTGACTGATCAAGCACCACAGATACAAATGGCCACTGCCAGACAATG- 1614            |
| 575  | Db | ysValLysAspGlnLeuLysValSerProPheLeuProAspAlaSerMetGluLysGluL 595   |
| 1615 | QY | -----CTAGTCACAGGCTCACCATCC 1636                                    |
| 595  | Db | eullePheAspGlyGlyLeuGlySerGlySerGlyGlnLysValAspLeulleThrTrpP 615   |
| 1637 | QY | CCACCAAGTATTTCTCGAATCCGCAATCGCTCGGGAATTTACATCCACCTGCAT 1696        |
| 615  | Db | roTrpSerGluThrSerSer--GluLysSerAlaGluProLeuSerLysProTrpLeuG 634    |
| 1697 | QY | CTTCAGATCAGACCGCATCAAGTCAGGTGCGCAAGAT- 1744                        |
| 634  | Db | luAspAspSerLeuLeuProAlaGluIleGluAspLysLysLeuValleuValAspL 654      |
| 1745 | QY | ACCTAGATCAAAATGGAT- 1761   |
| 654  | Db | ysMetAspSerThrAspGlnIleSerLysHisSerLysTyValHisAspAspArgSerI 674    |
| 1762 | QY | -----CTCTCTGACACTC 1774  |
| 674  | Db | leHisPheProGluGluGluProLeuSerGlyProAlaValProIlePheAlaAspThrA 694   |
| 1775 | QY | CTGCCCCCA- 1804  |
| 694  | Db | laAlaGluSerAlaSerLeuThrLeuProLysHisIleSerGluValProGlyValAspA 714   |
| 1805 | QY | AATATGTTTCTGTCCCATCATTTTCTTGAGGATACACCTCTCTGTCTCAGT- 1857          |
| 714  | Db | spTySerValThrLysAlaProLeulleuLeuThrSerValAlaIleSerAlaSerThrA 734   |
| 1858 | QY | -----TTACAGTATATCACCCTAGTT 1879                                    |

Db 734 sPlysserAaspGlnAlaAaspAlaLeuArgGluAaspMetGluGlnThrGluSerS 754  
 QY 1880 CT-----  
 Db 754 erAanTyrrGluTrpPheAaspSerGluValSerMetValLysProaspMetGlnThrLeuT 774  
 QY 1883 TGACCAATTCCTCCCAAGGCGGAGAG-----  
 Db 774 rPThrileLeuProGluSerGluArgValTrpThrArgThrSerSerLeuGluLysLeuS 794  
 QY 1908 -----  
 Db 794 erArgAaspileLeuAlaSerThrProGlnSerAlaAaspArgLeuTrpLeuSerValThrG 814  
 QY 1908 -----  
 Db 814 lnSerThrLysLeuProProThrThrIleSerThrLeuLeuGluAaspGluValIleMetG 834  
 QY 1908 -----  
 Db 834 lyValGlnAaspIleSerLeuGluLeuAaspArgIleGlyThrAaspTyrrGlnProGluG 854  
 QY 1908 -----  
 Db 854 lnValGlnGluAaspGlyLysValGlySerTyrrValGluMetSerThrSerValHiss 874  
 QY 1908 -----  
 Db 874 erThrGluMetValSerValAlaTrpProThrGluGlyAaspAspLeuSerTyrrThrG 894  
 QY 1909 -----CTGGTAGTTCCTTCAGTCTGGTGTCTTAACATGGCTCTCTCA 1954  
 Db 894 lnThrSerGlyAlaLeuValPhePheSerLeuArgValThrAsnMetMetPheSerG 914  
 QY 1955 AGCACTCTTCAACAGAGCTCTCTGGAGTACCGAGCTCTGGAGCAACAATTCACACAGC 2014  
 Db 914 luAaspLeuPheAsnLysAsnSerLeuGluTyrrLysAlaLeuGluGlnArgPheLeuGluL 934  
 QY 2015 TGCTGGTTCCTATACGATCCAACTTACAGGATTTAAGCACTTGAATTAACCTTAAC 2074  
 Db 934 euLeuValProTyrrLeuGlnSerAsnLeuThrGlyPheGlnAsnLeuGluLeuAsnPro 954  
 QY 2075 TCAGAAACGGAGGTGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2134  
 Db 954 heArgAsnGlySerIleValValAsnSerArgMetLysPheAlaAsnSerValProPro 974  
 QY 2135 ACTCACCAAGGCTGTGCACGGGCTCTTGGAGGATTTCTGCTGCTGAGCCCAACAC 2194  
 Db 974 snValAsnAlaValTyrrMetIleLeuGluAaspPheCysThrThrAlaTyrrAsnThrM 994  
 QY 2195 TCCATCTGGAATAGACAGCTACTCTCTCAACATTTGAACAGCTGATCAAGCAGATCCCT 2254  
 Db 994 etAsnLeuAlaIleAaspLysTyrrSerLeuAaspValGluSerGlyAaspGluAlaAsnPro 1014  
 QY 2255 GCAAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2314  
 Db 1014 yeLysPheGlnAlaCysAsnGluPheSerGluCysLeuValAsnProTrpSerGlyGlu 1034  
 QY 2315 CGAGTGTGCTGCCAAACAGGATATGACAGCAG-----GGAGCTCTGG 2359  
 Db 1034 laLysCysArgCysPheProGlyTyrrLeuSerValGluGluArgProCysGlnSerLeu 1054  
 QY 2360 AGGCTGTGGAACAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2419  
 Db 1054 yeAaspLeuGlnProAaspPheCysLeuAsnAaspGlyLys---CysAspIleMetProGlyH 1073  
 QY 2420 AGGAGGCTTCCATCGAGTGTGCAGATC-----ACTCTGAAATCAGCATACA 2467  
 Db 1073 isGlyAlaIleCysArg-CysArgValGlyGluAsnTrpTrpTyrrGlyGlyLysHicCys 1092  
 QY 2468 AAATAGTGTAAAGTTCCAAATCAACAAATTAACAGGTAATCAGTAAAGAAAT 2527  
 ...

Db 1093 GluGluPheValSerGluProValIleIleGlyIleThrIleAlaSerValValGlyLeu 1112  
 QY 2528 CTGAATTAATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATG 2587  
 Db 1113 Leu-----  
 QY 2588 AACTGAAATGTACAAATATCCTAGCTATCTCAAGAGAGATGATTTGGCTCTCTCAAG 2647  
 Db 1114 -----VallePheSerAlaIleIleTyrrPhePheIleArgThrLeuGln 1128  
 QY 2648 GAAATGGACAGCAGCAT-----ATTCAATGGTCAATCAAAATCCAGACAT 2692  
 Db 1129 AlaHisHisAaspArgSerGluArgGluSerProPheSerGlySerSerArgGlnProAasp 1148  
 QY 2693 ACAGTCAACACT---GAGAT-----CAGCACACACATATTTCAATATAGAAGA 2740  
 Db 1149 SerLeuSerSerIleGluAsnAlaValLysTyrrAsnProValTyrrGluSerHisArg 1167  
 RESULT 10  
 Q9UKD4  
 ID Q9UKD4 PRELIMINARY; PRT; 1241 AA.  
 AC Q9UKD4;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Interphotoreceptor matrix proteoglycan 200.  
 GN IMPG2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Retina;  
 RX MEDLINE=20391216; PubMed=10542133;  
 RA Kuehn M.H., Hageman G.S.;  
 RT "Molecular characterization and genomic mapping of human IPW 200, a  
 RL Mol. Cell Biol. Res. Commun. 2:1103-110(1999).  
 DR EMBL; AF173155; XAF06999.1; -;  
 DR Genew; HGNC:18362; IMPG2.  
 DR GO; GO:0004872; F:receptor activity; IEA.  
 DR InterPro; IPR006209; EGF like.  
 DR InterPro; IPR000082; SEA\_domain.  
 DR Pfam; PF01390; SEA; 2.  
 DR SMART; SM00200; SEA; 2.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS50024; SEA; 1.  
 KW Receptor.  
 SQ SEQUENCE 1241 AA; 138560 MW; 45831BFB848245D3 CRC64;  
 Alignment Scores:  
 Pred. No.: 2,54e-58 Length: 1241  
 Score: 917.00 Matches: 318  
 Percent Similarity: 39.23% Conservative: 161  
 Best Local Similarity: 26.04% Mismatches: 311  
 Query Match: 15.61% Indels: 431  
 DB: Gaps: 39  
 US-10-007-270-1 (1-3330) x Q9UKD4 (1-1241)  
 QY 170 ATTTTCTCCAAAGTTCAGGAACC---AAGATATCTCCATTACATATACCATCTGAA 226  
 Db 16 IlePheValLeuIleGluGlyAaspPheProSerLeuThrAlaGlnThrTyrrLeuSer--- 34  
 QY 227 ACTAAGACATACAGCAATCCC-----CCAGAATGAACA 262  
 Db 35 IleGluGluIleGlnGluProLysSerAlaValSerPheLeuLeuProGluSerThr 54  
 QY 263 -----ACTGAAAGTACTGAAAAAATGACAAATGTCACATATGACGCA 307  
 Db 55 AspLeuSerLeuAlaThrLysLysGlnProLeuAspArgArgGluThrGluArgGln 74

|      |     |         |         |      |        |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
|------|-----|---------|---------|------|--------|------|------|-----|------|------|------|------|-------|------|------|------|-----|-----|-----|-----|-----|
| 397  | al  | Trp     | Asn     | Thr  | Gln    | Ser  | Ser  | Leu | Gln  | Ala  | Thr  | Pro  | Ser   | Ser  | Ile  | Leu  | Asp | Asn | Thr | Pro | 417 |
|      |     |         |         |      | :::::  |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
| 1324 | QY  |         |         |      |        |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
|      |     |         |         |      |        |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
| 417  | he  | Gln     | Ala     | Ala  | Trp    | Pro  | Ser  | Ala | Asp  | Glu  | Ser  | Ile  | Thr   | Ser  | Ser  | Ile  | Pro | Leu | Asp | Pro | 437 |
|      |     |         |         |      | :::    |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
|      |     |         |         |      |        |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
| 1348 | QY  |         |         |      |        |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
|      |     |         |         |      |        |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
| 437  | he  | Ser     | Ser     | Gly  | Pro    | Pro  | Ser  | Ala | Thr  | Gly  | Arg  | Glu  | Ile   | Trp  | Ser  | Glu  | Ser | Pro | Leu | Gly | 457 |
|      |     |         |         |      |        |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
| 1380 | QY  |         |         |      |        |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
|      |     |         |         |      |        |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
| 457  | sp  | Leu     | Val     | Ser  | Thr    | His  | Lys  | Leu | Ala  | Phe  | Pro  | Ser  | Lys   | Met  | Gly  | Leu  | Ser | Ser | Ser | Pro | 477 |
|      |     |         |         |      |        |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
| 1381 | QY  |         |         |      |        |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
|      |     |         |         |      |        |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
| 477  | lu  | Val     | Leu     | Glu  | Val    | Ser  | Ser  | Leu | Thr  | Ile  | His  | Ser  | Val   | Thr  | Pro  | Ala  | Val | Leu | Gln | Thr | 497 |
|      |     |         |         |      |        |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
| 1385 | CTT | TGCTGTT | TATA    | ACAG | AGG    | TGCT | ACT  | TGT | TAGT | CCAG | AACT | CTCT | CTGTT | GAAC | CCCC | 1444 |     |     |     |     |     |
|      |     |         |         |      |        |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
| 497  | ly  | Leu     | Pro     | Val  | Ala    | Ser  | Glu  | Arg | Thr  | Ser  | Gly  | Ser  | His   | Leu  | Val  | Glu  | Asp | G   | 515 |     |     |
|      |     |         |         |      |        |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
| 1445 | AGC | TTG     | ACAG    | CAGT | GGC    | AGC  | AGAG |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
|      |     |         |         |      |        |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
| 515  | ly  | Leu     | Ala     | Asn  | Val    | Glu  | Glu  | Ser | Glu  | Asp  | Phe  | Leu  | Ser   | Ile  | Asp  | Ser  | Leu | Pro | Ser | Ser | 535 |
|      |     |         |         |      |        |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
| 1490 | CTT | GTCTCC  | ACCTGCT | ATG  | CGCTCT | AC   |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
|      |     |         |         |      |        |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
| 535  | er  | Phe     | Thr     | Gln  | P      | ro   | Val  | Pro | Lys  | Glu  | Thr  | Ile  | Pro   | Ser  | Met  | Glu  | Asp | Ser | Val | Ser | 555 |
|      |     |         |         |      |        |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
| 1523 | TTC | ACAG    | AGT     | TC   | CA     |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
|      |     |         |         |      |        |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
| 555  | eu  | Thr     | Ser     | Ser  | Pro    | Tyr  | Ile  | Thr | Ser  | Ser  | Ile  | Pro  | Phe   | Gly  | Leu  | Asp  | Ser | Leu | Thr | Ser | 575 |
|      |     |         |         |      |        |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |
| 1565 | CTC | TGCTGAT | CAAGC   | CAC  | CA     | CAT  | CA   | TAT | GC   | CA   | TG   | CC   | CA    | TG   | CA   | CAC  | AG  | CA  | ATG |     |     |
|      |     |         |         |      |        |      |      |     |      |      |      |      |       |      |      |      |     |     |     |     |     |

Db 754 erAsnTyrGluTrpPheAspSerGluValSerMetMetValIysProAspMetGlnThrLeuT 774  
 QY 1883 TGACCAATGCCCCCAAGGCCGAGAG----- 1908  
 Db 774 rPThrIleLeuProGluSerGluArgValTrpThrArgThrSerSerIleuGluIysLeuS 794  
 QY 1908 ----- 1908  
 Db 794 erArgAspIleLeuAlaSerThrProGlnSerAlaAspArgLeuTrpIleuSerValThrG 814  
 QY 1908 ----- 1908  
 Db 814 lnSerThrIysLeuProProThrThrIleSerThrIleuLeuGluAspGluValIleMetG 834  
 QY 1908 ----- 1908  
 Db 834 lyValGlnAspIleSerLeuGluLeuAspArgIleGlyThrAspTyrTyrGlnProGluG 854  
 QY 1908 ----- 1908  
 Db 854 lnValGlnGluGlnAsnGlyLysValGlySerTyrValGluMetSerThrSerValHiss 874  
 QY 1908 ----- 1908  
 Db 874 erThrGluMetValSerValAlaTrpProThrGluGlyAspAspLeuSerTyrThrG 894  
 QY 1909 -----CTGCTAGTCTTCTTCAGTCTGGTGTCTTCAACATGCGCTCTCTCA 1954  
 Db 894 lnThrSerGlyAlaLeuValPhePheSerLeuArgValThrAsnMetMetPheSerG 914  
 QY 1955 ACGACTCTTCAACAAGAGCTCTCTGGAGTACCGAGCTCTGGAGCAACAATTCACACAGC 2014  
 Db 914 luAspLeuPheAsnLysAsnSerLeuGluTyrIysAlaLeuGluGlnArgPheLeuGlu 934  
 QY 2015 TGCTGGTTCCTATCTACGATCCAACTTACAGATTAAAGCACTTGAATTAATACT 2074  
 Db 934 euLeuValProTyrLeuGlnSerAsnLeuThrGlyPheGlnAsnLeuGluIleLeuAsn 954  
 QY 2075 TCAGAAACGGAGGTGATTGTGAATAGCAAAATGAGTTTCTGCTGCTGCGGTATA 2134  
 Db 954 heArgAsnGlySerIleValValAsnSerArgMetIysPheAlaAsnSerValProPro 974  
 QY 2135 ACCTCACAAGGCTGTGCGGCGTCTGAGGATTTCTGCTGCTGCGAGCCCAACAC 2194  
 Db 974 snValAsnAsnAlaValTyrMetIleLeuGluAspPheCysThrAlaTyrAsnThrM 994  
 QY 2195 TCATCTGGAATAGACAGCTACTCTCTCAACATTGAACCGAGCTGATCAAGCAGATCCCT 2254  
 Db 994 etAsnLeuAlaIleAspIysTyrSerLeuAspValGluSerGlyAspGluAlaAsnProC 1014  
 QY 2255 GCAAGTCTCTGCGCTCGCGCAATTTGCCCAATGTGTAAGCAACGACGAGCTGAGGAG 2314  
 Db 1014 yslYsPheGlnAlaCysAsnGluPheSerGluCysLeuValAsnProTrpSerGlyGluA 1034  
 QY 2315 CGAGTCTGCTGCAAAACGAGATATACAGCCAG-----GGGAGCTCTGG 2359  
 Db 1034 laLysCysArgCysPheProGlyTyrLeuSerValGluGluArgProCysGlnSerLeu 1054  
 QY 2360 ACGTCTGGAACAGCCCTCTGTGGCCCTGGCCAAAGCAAGATGAGGCTCTCCAGGAA 2419  
 Db 1054 yAspLeuGlnProAspPheCysLeuAsnAspGlyLys--CysAspIleMetProGlyH 1073  
 QY 2420 AGGAGTCTCCATCCAGTGTGCAGATC-----ACTCTGAAATCAGCATACA 2467  
 Db 1073 iglyAlaIleCysArg-CysArgValGlyGluAsnTrpTrpTyrArgGlyIysHicCys 1092  
 QY 2468 AAATAGTGTAAAAATGTTCCAAATCAACAAATAACAGGTAATCAGTAAAGAAAT 2527  
 Db 1093 GluGluPheValSerGluProValIleIleGlyIleThrIleAlaSerValGlyLeu 1112  
 QY 2528 CTGAATTACTACCGGTAGATATGAAGATTTAACAATCAGATTCGGGAAGAAATTA 2587  
 Db 1113 Leu----- 1113

QY 2588 AACTGAAATGTACAAATTATCACTTAGCTATCTCAAGAGAGATGATTTGCTTCTCAAG 2647  
 Db 1114 -----ValIlePheSerAlaIleIleTyrPhePheIleArgThrLeuGln 1128  
 QY 2648 GAAATGTGAGACAGCAT-----ATTCATGGTTCATCAAAATCCAGACAT 2692  
 Db 1129 AlahIshIAspArgSerGluArgGluSerProPheSerGlySerSerArgGlnProAsp 1148  
 QY 2693 ACAGTCAACACT--GAGAT-----CAGCACACACATATTTCAATATAGAGA 2740  
 Db 1149 SerLeuSerSerIleGluAsnAlaValIysTyrAsnProValTyrGluSerHisArg 1167  
 RESULT 11  
 Q9BZV3  
 ID Q9BZV3 PRELIMINARY; PRT; 1241 AA.  
 AC Q9BZV3;  
 DT 01-JUN-2001 (T-EMBLrel. 17, Created)  
 DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)  
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)  
 DE Interphotoreceptor matrix proteoglycan 200.  
 GN IMPG2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20391216; PubMed=10542133;  
 RA Kuehn M.H., Hageman G.S.;  
 RT "Molecular characterization and genomic mapping of human IPM 200, a  
 second member of a novel family of proteoglycans.";  
 RL Mol. Cell Biol. Res. Commun. 2:103-110(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Kuehn M.H., Hageman G.S.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF271379; AAG49889.1; JOINED.  
 DR EMBL; AF271363; AAG49889.1; JOINED.  
 DR EMBL; AF271364; AAG49889.1; JOINED.  
 DR EMBL; AF271365; AAG49889.1; JOINED.  
 DR EMBL; AF271366; AAG49889.1; JOINED.  
 DR EMBL; AF271367; AAG49889.1; JOINED.  
 DR EMBL; AF271368; AAG49889.1; JOINED.  
 DR EMBL; AF271369; AAG49889.1; JOINED.  
 DR EMBL; AF271370; AAG49889.1; JOINED.  
 DR EMBL; AF271371; AAG49889.1; JOINED.  
 DR EMBL; AF271372; AAG49889.1; JOINED.  
 DR EMBL; AF271373; AAG49889.1; JOINED.  
 DR EMBL; AF271374; AAG49889.1; JOINED.  
 DR EMBL; AF271375; AAG49889.1; JOINED.  
 DR EMBL; AF271376; AAG49889.1; JOINED.  
 DR EMBL; AF271377; AAG49889.1; JOINED.  
 DR EMBL; AF271378; AAG49889.1; JOINED.  
 DR GO; GO:0004872; F:receptor activity; IEA.  
 DR InterPro; IPR006209; EGF-like.  
 DR InterPro; IPR000082; SEA\_domain.  
 DR Pfam; PF01390; SEA; 2.  
 DR SMART; SM00200; SEA; 2.  
 DR PROSITE; PS01186; EGF 2; 1.  
 DR PROSITE; PS50024; SEA; 1.  
 KW Receptor.  
 SQ SEQUENCE 1241 AA; 138628 MW; 836208513456B2B CRC64;  
 Alignment Scores:  
 Pred. No.: 3,01e-58 Length: 1241  
 Score: 916.00 Matches: 315  
 Percent Similarity: 39.07% Conservative: 162  
 Best Local Similarity: 25.80% Mismatches: 313  
 Query Match: 15.59% Indels: 431  
 DB: 4 Gaps: 38  
 US-10-007-270-1 (1-3330) x Q9BZV3 (1-1241)



|    |      |   |      |
|----|------|---|------|
| Db | 694  | laAlaGluSerAlaSerLeuThrLeuProLysHisIleSerGluValProGlyValAspA  | 714  |
| QY | 1805 | AATATGTTCTTCTCCACAGATCATTTCTTGAGAGTACACACTCTCTGCTCAGCT        | 1857 |
| Db | 714  | spTyzSerValThrLysAlaProLeuIleLeuThrSerValAlaIleSerAlaSerThrA  | 734  |
| QY | 1858 | -----TTACAGTATATCACCACCTAGTT                                  | 1879 |
| Db | 734  | spLysSerAspGlnAlaAspAlaIleLeuArgGluAspMetGluGlnIleThrGluSerS  | 754  |
| QY | 1880 | CT-----A  | 1882 |
| Db | 754  | erAsnTyzGluTrpPheAspSerGluValSerMetValLysProAspMetGlnThrLeuT  | 774  |
| QY | 1883 | TGACCATTCGCCCCCAAGGGCCGAGAG                                   | 1908 |
| Db | 774  | rphThrIleLeuProGluSerGluArgValTrpThrArgThrSerSerLeuGluLysLeuS | 794  |
| QY | 1908 | -----   | 1908 |
| Db | 794  | erArgAspIleLeuAlaSerThrProGlnSerAlaAspArgLeuTrpLeuSerValThrG  | 814  |
| QY | 1908 | -----   | 1908 |
| Db | 814  | lnSerThrLysLeuProProThrThrIleSerThrLeuLeuGluAspGluValIleMetG  | 834  |
| QY | 1908 | -----   | 1908 |
| Db | 834  | lyValGlnAspIleSerLeuGluLeuAspArgIleGlyThrAspTyzTyzGlnProGluG  | 854  |
| QY | 1908 | -----   | 1908 |
| Db | 854  | lnValGlnGluGlnAsnGlyLysValGlySerTyzValGluMetSerThrSerValHisS  | 874  |
| QY | 1908 | -----   | 1908 |
| Db | 874  | erThrGluMetValSerValAlaTrpProThrGluGlyAspAspLeuSerTyzThrG     | 894  |
| QY | 1909 | -----CTGTGTAGTGTCTTCAGTCTGCTGTCTTAACATGGCTCTCTCCCA            | 1954 |
| Db | 894  | lnThr-SerGlyAlaLeuValValPhePheSerLeuArgValThrAsnMetMetPheSerG | 914  |
| QY | 1955 | ACGACCTGTTACACAGAGCTCTCTGGAGTACCGAGCTCTGGAGCAACAATTCACACAGC   | 2014 |
| Db | 914  | luAspLeuPheAsnLysAsnSerLeuGluTyzLysAlaLeuGluGlnArgPheLeuGluL  | 934  |
| QY | 2015 | TGCTGTCTTCATATCTACAGTCCAACTTACAGATTTAAAGCACTTGAATTACTTA       | 2074 |
| Db | 934  | euLeuValProTyzLeuGlnSerAsnLeuThrGlyPheGlnAsnLeuGluIleLeuAsn   | 954  |
| QY | 2075 | TCAGAAACGGGAGTGTGATTGTGAATAGCAAAATGAAGTTTCTTAAGTCTGTGCCGTATA  | 2134 |
| Db | 954  | heArgAsnGlySerIleValValAsnSerArgMetLysPheAlaAsnSerValProProA  | 974  |
| QY | 2135 | ACCTCACCAAGGCTGTGACCGGGTCTTGGAGGATTTTCGTTCTGTGCGCCCAACAAC     | 2194 |
| Db | 974  | snValAsnAsnAlaValTyzMetIleLeuGluAspPheCysThrThrAlaTyzAsnThrM  | 994  |
| QY | 2195 | TCCATCTGGAATATAGACAGCTACTCTCTCAACTGACACGAGTATCAAGCAGATCCCT    | 2254 |
| Db | 994  | erAsnLeuAlaIleAspLysTyzSerLeuAspValGluSerGlyAspGluAlaThrProC  | 1014 |
| QY | 2255 | GCAAGTCTCGGCGTGGCGGAATTTGGCCCAATGTGTAAAGAACGACGACTGGAGAG      | 2314 |
| Db | 1014 | ylsLysPheGlnAlaCysAsnGluPheSerGluCysLeuValAsnProTrpSerGlyGluA | 1034 |
| QY | 2315 | CGGAGTGTCTGCTCCAAACACGAGATATGACAGCCAG                         | 2359 |
| Db | 1034 | lalyeCysArgCysPheProGlyTyzLeuSerValGluGluArgProCysGlnSerLeu   | 1054 |
| QY | 2360 | ACGGTCTGGAAACGAGCCCTCTGTGGCCCTGGCACAAGGAATGCGAGGTCTCCACGGAA   | 2419 |

1054 yAspPleuGlnProAspPheCysYleuAenAspGlyLys---CysAspPleMetProGlyH 1073  
 QY 2420 AGGAGCTCCATGCAGCTGGTCGACATC-----ACTCTGAAATCAAGCATACA 2467  
 Db 1073 isGlyAlaIleCysArg-CysArgValGlyGluAenTptTpyrGlyLysHisCys 1092  
 QY 2468 AACTAGTGTGTTAAAGTTCCAAATCAACAAATAACAAGGTAATCAGTAAAGAAAT 2527  
 Db 1093 GluGluPheValSerGluProValIleGlyIleThrIleAlaSerValValGlyLeu 1112  
 QY 2528 CTGAATTACTGACCGTAGAATATGAAGATTTAACCATCAAGATTCGGAGGAATTTAA 2587  
 Db 1113 Leu----- 1113  
 QY 2588 AACTGAAATGTACAAATTATCATCTTAGCTATCTCAAGAGAGATGATTGCCTTCTCAAG 2647  
 Db 1114 -----ValIlePheSerAlaIleIleTyPhePheIleArgThrIleuGln 1128  
 QY 2648 GAAATGCGACAGGCAT-----ATTCATGGGTCATCAAAATCCAGACAT 2692  
 Db 1129 AlaHisAspArgSerGluArgGluSerProPheSerGlySerSerArgGlnProAsp 1148  
 QY 2693 ACAGTCAACACT---GAGAAT-----CAGCACACACCATATTCAAATATAGAAGA 2740  
 Db 1149 SerLeuSerSerIleGluAsnAlaValIleTyAsnProValTyGluSerHisArg 1167  
 RESULT 12  
 Q810Y3 PRELIMINARY; PRT; 1127 AA.  
 AC Q810Y3;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Sialoprotein associated with cones and rods proteoglycan.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22477978; PubMed=12589770;  
 RA Chen Q., Lee J.W., Nishiyama K., Shadrach K.G., Rayborn M.E.,  
 RA Hollyfield J.G.;  
 RT "SPACRAN in the interphotoreceptor matrix of the mouse retina:  
 RT molecular, developmental and promoter analysis.";  
 RL Exp. Eye Res. 76:1-14 (2003).  
 DR EMBL; AY174061; AAC02121.1; -;  
 DR InterPro; IPR006209; EGF like  
 DR InterPro; IPR000082; SEA\_domain.  
 DR Pfam; PF00008; EGF; 1.  
 DR Pfam; PF01390; SEA; 2.  
 DR SMART; SM00200; SEA; 2.  
 DR PROSITE; PS01186; EGF 2; 1.  
 DR PROSITE; PS50024; SEA; 2.  
 SQ SEQUENCE 1127 AA; 125812 MW; B51B3DBE69606708 CRC64;  
 Alignment Scores:  
 Pred. No.: 1.97e-56 Length: 1127  
 Score: 891.00 Matches: 278  
 Percent Similarity: 42.01% Conservative: 140  
 Best Local Similarity: 27.94% Mismatches: 273  
 Query Match: 15.17% Indels: 305  
 DB: 11 Gaps: 35  
 US-10-007-270-1 (1-3330) x Q810Y3 (1-1127)

[illegible]



QY 329 -----CGAACAAGATCCGCAATTTTCCCAAGGGGGTAAAGTCGT 373  
 Db 64 ArgArgTrpLeuLeuArgArgArgSerIleLeuPheProAsnGlyValIleCys 83  
 QY 374 CCACAGGAATCCATGAACACAGATTATTAGACAGTCTTCAAGCTTATTATAGATTGAGATG 433  
 Db 84 SerSerGluThrValAlaGluAlaValAlaAsnHisValLysTyrPheLysAlaArgVal 103  
 QY 434 TGTCCAGGAAGCAGTATGGGAACATATCGGATCTTCTTGGATCCGATCCCTCCAGACAGG 493  
 Db 104 CysGlnGluAlaIleTrpGluAlaPheArgThrPheTrpAspArgLeuProGlyArgAsp 123  
 QY 494 GAATATCAGGACTGGGTTCAGCATCTGCCAGCAGAGACCTTCTCCCTTTGACATTCGA 553  
 Db 124 GluTyrArgHisTrpMetAsnLeuCysGluAspGlyValThrSerValPheGluMetGly 143  
 QY 554 AAAAACTTCAGCAATCCCGAGGACCTCGATCTTCTCCAGCAGAGATA----- 604  
 Db 144 AlaHisPheSerGlnSerValGluHisArgAsnLeuIleMetLysLysLeuAlaTyrThr 163  
 QY 605 AAACAGAGAAGTTTCCCTCGACAGAAAGATCAATAATCTGCAGAGAGACATTCGGAGAG 664  
 Db 164 ArgGluAlaGluSerSerSerCysLysAspGln---SerCysGlyProGluLeuSerPhe 182  
 QY 665 CCT-----GGTGAACCAATTGTCATTTCAACAGCAATCTACATTCAAAGACTTGG 715  
 Db 183 ProValProIleGlyGluThrSerThrLeuThrGlyAlaVal----- 196  
 QY 716 CGAGTATTTCAAGAAACCTCAGAGAGCAAAATTCAGAGTGTGCCACGCTCCTCCTTG 775  
 Db 197 -----SerSerAlaSerTyrPro 202  
 QY 776 GGCCTTCCTCCT-----CTCCTCTCTGTGACACCTCTCTCTCAATCAATTCCTC 822  
 Db 203 GlyLeuAlaSerGluSerSerAlaAlaSerProGlnGluSerIleSerAsnGluIle--- 221  
 QY 823 GATATACATCTAACACACCAAGATCCCTACACAGAAAGAGAAACAGATTCCTGCTGTG 882  
 Db 222 GluAsnValThrGluGluProThrGlnProAlaAlaGluGlnIleAlaGluPhe----- 239  
 QY 883 TTGGAGGAGCAGAGGGTGGAGTCTCTCTCTGTGTAACACAGAGTTCACAGCAGAG 942  
 Db 240 -----SerIleGlnLeuLeuGlyLysArgTyrSerGluGlu 251  
 QY 943 CTGCGTCACTCCAGTCCCATATTACAGAGAGTACAGAGAAAGTCCCACTTCAGATG 1002  
 Db 252 LeuArgAspProSerSerAlaLeuTyrArgLeuLeuValGluGluPheIleSerGluVal 271  
 QY 1003 CAAGAATATTAAAGAACTCCAGGATTCAAAAAATCCATGTGTAGGATTAGACCA 1062  
 Db 272 GluLysAlaPheThrGlyLeuProGlyTyrLysGlyIleArgValLeuGluPheArgAla 291  
 QY 1063 AAGAAAGAAAGATGGTCAAGCTCCACAGAGATGCAACTTACGGCCCTTTTAAGAGA 1122  
 Db 292 ProGluGluAsnAspSerGlyLeuValHisTyrAlaValThr----- 306  
 QY 1123 CACAGTCAGAGCAAAAGCCCTGCAAGT---GACCTCCCTGCTTTTGTATCCAAACAA 1179  
 Db 307 PheAsnGlyGluAlaIleSerAsnThrThrTrpAspLeuIleSerLeuHisSerAsnLys 326  
 QY 1180 ATTCAAGTCAGGAAGTCTATCATGCAACCATGGAGGAGCAGACCAACCAATCTAT 1239  
 Db 327 ValGluAsn-----HisGlyLeuValGluMetAspLysProThrAlaVal 342  
 QY 1240 CTCACAGCTACAGACCTCAAAAGGCTGATCAGCAAGCACTAGAGGAAGAAATCTTTG 1299  
 Db 343 TyrThrIleSerAsnPheArgAspTyrIleAlaGluThrLeu-----HisGlnAsnPhe 360  
 QY 1300 GATGTGGGGCAATTCAGTTCATCGATGAATTCGTGATCATCTGCCACCTTTGGTCT 1359  
 Db 361 LeuMetGlyAsnSerSerLeuAsnProAsp---ProLysSerLeuGluLeuIleAsnAla 379

QY 1360 GACACCCATCAGAGCTGCC-----ACATCTTTGCTGTGTATAACA 1401  
 Db 380 ValLeuGlnProAspLeuProValAlaProGluGlyArgThrSerGlySerPheIleLeu 399  
 QY 1402 GAGATGCTACTTGTGATCCAGAA----- 1425  
 Db 400 GluAspGlyLeuAlaSerThrGluGluLeuGluAspThrSerIleAspGlyLeuProSer 419  
 QY 1426 -----CTTCTCTCTGTGTGA----- 1440  
 Db 420 SerProLeuIleGlnProValProLysGluThrValProProMetGluAspSerAspThr 439  
 QY 1441 -----CCCCAGCTTCAG-----ACAGTGGAC--- 1461  
 Db 440 AlaLeuLeuSerThrProHisLeuThrSerSerAlaIleGluAspLeuThrLysAspIle 459  
 QY 1462 GGAGCAGAGCATGCTCTA----- 1479  
 Db 460 GlyThrProSerGlyLeuGluSerLeuAlaSerAsnIleSerAspGlnLeuGluValIle 479  
 QY 1480 -----CCTGACACTTCT----- 1491  
 Db 480 ProTrpPheProAspThrSerValGluLysAspPheIlePheGluSerGlyLeuGlySer 499  
 QY 1492 -----TGCTCTCCACCTGCTATGGCTCT 1515  
 Db 500 GlySerGlyLysAspValAlaAspValIleAspTrpProTrpSerGluThrSerLeuGluLys 519  
 QY 1516 ACC-----TCCCTGTCAAGACTCCA----- 1536  
 Db 520 ThrThrLysProLeuSerLysSerTrpSerGluGluGlnAlaLeuLeuProThrGlu 539  
 QY 1536 ----- 1536  
 Db 540 GlyArgGluLysLeuHisIleAspGlyArgValAspSerThrGluGlnIleIleGluSer 559  
 QY 1536 ----- 1536  
 Db 560 SerGluHisArgTyrGlyAspArgProIleHisPheIleGluGluGluSerHisValArg 579  
 QY 1537 -----CCTTCTTTATGGCATCAGC-----ATCTTCTCT--- 1566  
 Db 580 SerThrIlePheIlePheValGluSerAlaThrProProThrSerProIlePheSerLys 599  
 QY 1567 -----CTGACTGATCAAGCCACACAGATACATGCGCATCGACCAATGCTAGTACCA 1623  
 Db 600 HisThrSerAspValProAspIleAspSerTyrSerLeuThrLysProPheLeuPro 619  
 QY 1624 -----GGGCTCACCATCCCCACACAGTATTTCTGCATCAGCAACTGGCTCTGGGAAT 1680  
 Db 620 ValThrIleAlaIleProAlaSerThrLysThrAspGluValLeuLysGluAspMet 639  
 QY 1681 TCACATCCACCTGCATCTTTCAGAT-----GACAGCCGATCAAGTGCAGTGGCGAA 1731  
 Db 640 ValHisThrGluSerSerHisLysGluLeuAspSerGluValProValSerArgPro 659  
 QY 1732 GATATG----- 1737  
 Db 660 AspMetGlnProValTrpThrMetLeuProGluSerAspThrValTrpThrArgThrSer 679  
 QY 1738 -----GTCAGACACCTAGATGAATGGTCTCTCTGACACTCTCT--- 1776  
 Db 680 SerLeuGlyLysLeuSerArgAspThrLeuAlaSerThrProGluSerThrAspArgLeu 699  
 QY 1777 -----GCCCCATCTGAGGTACCA-----GAG 1797  
 Db 700 TrpLeuLysAlaSerMetThrGlnSerThrGluLeuProSerThrThrHisSerThrGln 719  
 QY 1798 CTCACGGAATATGTTTCTGTC----- 1818  
 Db 720 LeuGluGluGluValIleMetAlaValGlnAspIleSerLeuGluLeuAspGlnValGly 739  
 QY 1819 -----CCAGATCATTTCTTG 1833

```

Db 740 ThrAspTyrTyrGlnSerGluLeuThrGluGlnHisGlyLeuAlaAspSerTyrVal 759
QY 1834 GAGATACCACTCTGCTCAGCTTTACAGTATATCACCAGTATCTTACGACCATGGC 1893
Db 760 GluMetSerThrSerVal-----HisTyrThrGluMetProIleValAlaLeuPro 776
QY 1894 CCCAAGGCC-----CGAGAGCTGTGTAGTCTTCTTCAT 1926
Db 777 ThrLysGlyGlyValLeuSerHisThrGlnThrAlaGlyAlaLeuValPhePheSer 796
QY 1927 CTGCGTGTGCTAATCATGCTCTCCAAACAGCTGTTCACACAGAGCTCTCTGGAGTAC 1986
Db 797 LeuArgValThrAsnMetLeuPheSerGluAspLeuPheAsnLysSerLeuGluTyr 816
QY 1987 CGAGCTCTGGAGCAACAAATTCACAGCTGTGTGTTCCATATCTACGATCCATCTTACA 2046
Db 817 LysAlaLeuGluGlnArgPheLeuGluLeuValProTyrLeuGlnSerAsnLeuSer 836
QY 2047 GCATTTAAGCAACTTGAATACTTACTTCAAGAGCGGTGTGATTTGGATAGACAA 2106
Db 837 GlyPheGlnAsnLeuGluIleLeuSerPheArgAsnGlySerIleValValAsnSerArg 856
QY 2107 ATGAGTTTGTAAAGTCTGTGCGGTATACCTCACCAGGCTGTGCAGCGGTCTTGGAG 2166
Db 857 ValArgPheAlaGluSerAlaProProAsnValAsnLysAlaMetTyrArgIleLeuGlu 876
QY 2167 GATTTCGTCTGTGAGCCCAACACTCCATCTGGAATAGACAGCTACTCTCTCAAC 2226
Db 877 AspPheCysThrThrAlaTyrGlnThrMetAsnLeuAspIleAspLysTyrSerLeuAsp 896
QY 2227 ATTGAACACAGCTGTCAAGCAGATCCCTGCAAGTCTCTGCGCTGCGGGAATTTGCCAA 2286
Db 897 ValGluSerGlyAspGluAlaAsnProCysLysPheGlnAlaCysAsnGluPheSerGlu 916
QY 2287 TGTGTAAGAAGCAACCGACTGAGGAGCGAGTGTGCTGCACACAGCATATGACAGC 2346
Db 917 CysLeuValAsnProTyrSerGlyGluAlaLysCysLysCysTyrProGlyTyrLeuSer 936
QY 2347 CAGGGG-----AGCTGACGCTGTGAACACGAGCTCTGTGCGCTCTGCGCTGCGC 2391
Db 937 ValAspGluLeuProCysGlnSerLeuCysAspLeuGlnProAspPheCysLeuAsnAsp 956
QY 2392 ACAAGGAATGCGAGGTCTCCAGGAGGAGGAGGAGCTCCATGCGAG 2436
Db 957 GlyLys---CysAspIleMetProGlyHisGlyAlaIleCysArg 970

RESULT 13
P70628
ID P70628 PRELIMINARY; PRT; 1239 AA.
AC P70628;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE FG10.2 protein.
GN FG10.2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCB1_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97038366; PubMed=8883960;
RA Wang X., Brownstein M.J., Young W.S. III;
RT "Sequence analysis of FG10.2, a gene expressed in the pineal gland and
RL the outer nuclear layer of the retina.";
DR EMBL; U76717; AAC52891.1;
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR00082; SEA_domain.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF01390; SEA; 2.

```

```

DR SMART; SM00181; EGF; 1.
DR SMART; SM00200; SEA; 2.
DR PROSITE; PS01186; EGF; 2; 1.
DR PROSITE; PS00024; SEA; 2.
SQ SEQUENCE 1239 AA; 137302 MW; 78CB5AE7A73E9B6 CRC64;

Alignment Scores:
Pred. No.: 3,78e-54 Length: 1239
Score: 860.00 Matches: 283
Percent Similarity: 37.59% Conservative: 129
Best Local Similarity: 25.82% Mismatches: 268
Query Match: 14.64% Indels: 416
DB: 11 Gaps: 34

US-10-007-270-1 (1-3330) x P70628 (1-1239)
QY 224 GAACTAAGACATAGACATATCCCAAGCA-----AATGAAACAACTGAAGTACTGAAAAA 280
Db 43 GluSerThrAspLeuSerLeuProThrArgLysArgGlnLeuLeuAspAlaThrGlu--- 61
QY 281 ATGTACAAAATGTCACTATGACAGCAATATTCGATTTCGCAAGCATGCAACAAAGA 340
Db 62 -----ThrGlyArgArg---TyrProLeu-----ArgArgArgArg 72
QY 341 TCCGCAATTTTCCCAACGCGGGTTAAAGTCTGTCCACAGGAATCCATGAAACAGATTTTA 400
Db 73 SerIleLeuPheProAsnGlyValArgLysCysProSerAspThrValAlaGluAlaVal 92
QY 401 GACAGCTTTCAGAGCTTATATGATGAGAGTGTTCAGGAGCAGTATGGAGACATAT 460
Db 93 AlaAsnHisValLysTyrPheLysAlaArgValCysGlnGluAlaIleTyrGluAlaPhe 112
QY 461 CGGATCTTTCTGATCGCATCCCTGCACAGCGGAATATCAGGAGTGGGTGAGCATCTGC 520
Db 113 ArgThrPheTyrAspArgLeuProGlyArgGluGluTyrGlnTyrTyrMetAsnLeuCys 132
QY 521 CAGCAGAGACCTCTGCTCTTTGACATTTGGAATAAACTTCAGCAATTCCTGAGGAGCAC 580
Db 133 GluAspGlyValThrSerValPheGluMetGlyThrGlnPheSerGlnSerValGluHis 152
QY 581 CTGATCTTCTCCAGCAGAGCAATA-----AAACAGAGAGTTTCCCTGCAGAGAA 631
Db 153 ArgHisLeuIleMetGluLysLeuThrTyrThrLysGluAlaGluSerSerCysLys 172
QY 632 GAT-----GAAATATCTGCAGAGAGACATTTGGAGAGCCTGTGTGAAC 676
Db 173 AspGlnAlaCysGlyProGluLeuSerSerProValProIleGlyLeuThrLeu 192
QY 677 ATT---GTCAATTCACAGCAATCTACATTTCAAGAGCTTGGGAGCTTCTAAGAAAC 733
Db 193 AlaGlyAlaValSerSerAlaSerTyr----- 201
QY 734 CCTCAGAGAGCAAAATCAAGATGTTGCAACGCTCTCACTTGGGCTTTTCCCTCTCACTC 793
Db 202 Pro-GlyAlaLaserGluArgSer-Ala-----AlaSerP 213
QY 794 CTGATGACACCTCTCTCAATGAAATCTCGATATATACATCAACACAGAGATGCTTA 853
Db 213 roGlnGluSerIleSerAsnGluIle-----GluAsnValThrGluGlnProT 229
QY 854 CAACAGAGAGAGAAACAGATTCGCTGTGTTGGAGCAGAGGGGTGAGCTCAGGCTCT 913
Db 229 hrProPro-----AlaAlaGluGlnIleAlaGluPheSerIleG 242
QY 914 CTCTGTGTAACAGAGAGTTCAGGAGCAGAGCTCGCTCACTCCAGTCCCACTATTACAGG 973
Db 242 LnLeuLeuGlyLysGlnTyrSerGluGluLeuArgAspProSerSerAlaLeuTyrArgL 262
QY 974 AGTACGAGGAAAGTCCCACTTTCAGATGCAAAAGATATTTAAGAAATTTCCAGGATCA 1033
Db 262 euLeuValGluGluPheIleSerGluValGluLysAlaPheThrGlyLeuProGlyTyrL 282
QY 1034 AAAAATCCATGTGTTAGGATTTAGACCCAAAGAAAGAAAGATGGCTCAGCTCCACAG 1093

```

```

Db      282 ysGlyIleHisValLeuAspPheArgSerProLysGluAsnGlySerGlyIleAspValH 302
QY      1094 AGATGCAACTTACGGCCATCTTTAAGACACACAGTCGACAGCAAGCAAAAGCCCTGCAAGT- 1152
Db      302 isTyRAlaValThr-----PheAsnGlyGluAlaIleSerAsnThrThr 317
QY      1153 --GACTCTCTCTTTGATTCCACAAATTAAGTGGAGGAGTCTATCATGGAACCA 1210
Db      317 rpAspLeuIleSerLeuHisAsnLysValGluAsn-----HisGlyLeuV 333
QY      1211 TGGAGGAGCAGCAAGCAACCAAGAAATCTATCTCACAGCTACAGACCTCAAAAGGCTGATCA 1270
Db      333 alGluLeuAspLysProThrAlaValTyThrIleSerAsnPheArgAspTyRileA 353
QY      1271 GCAAGCACTAGAG----- 1284
Db      353 laGluThrLeuHisGlnAsnPheLeuMetGlyAsnSerSerLeuAsnProAspProLysS 373
QY      1284 ----- 1284
Db      373 erLeuGlnLeuIleAsnValArgGlyValLeuLeuProGlnThrGluIleValTrpA 393
QY      1285 ----GAAGACAATCTTTGGATGGGGCAATTCAGTTCTACTCATGAAT- 1332
Db      393 snThrGlnSerSerLeuGlnValThrSerSerIleLeuAspAsnThrLeuGlnA 413
QY      1332 ----- 1332
Db      413 laGluTrpLeuSerAlaAspGluSerIleThrThrThrThrThrIleSerProp 433
QY      1333 -----GCTGGATCAGCCAGCCTTGGTCTGCTACACCCCAATCAGAG- 1374
Db      433 heGlyPheSerSerGlyProProSerAlaThrGlyArgGluLeuHisSerGluSerThrL 453
QY      1375 -----CTGCCACATCTTT- 1389
Db      453 euGlyAspIleValSerThrProLysLeuAlaSerProSerLysValValLeuSerSers 473
QY      1390 -----GCTGTTATAA 1399
Db      473 exProGluValLeuGlyGlySerSerLeuThrLeuHisSerValThrProAlaValLeuG 493
QY      1400 CAGAGATGCTACTTTGAGTCCAGAA----- 1425
Db      493 lnIleAspLeuProValAlaProGluGlyArgThrSerGlySerSerIleLeuGluAspA 513
QY      1425 ----- 1425
Db      513 spAsnThrGluGluSerGluAspValSerIleAspValLeuProSerSerSerLeuIleG 533
QY      1426 -----CTTCTCTCTTGA- 1440
Db      533 lnProValProLysGluThrValProMetGluAspSerAspMetIleLeuLeuThrs 553
QY      1441 -----CCCCACCTTGACAGTGG 1459
Db      553 erSerProHisLeuThrSerSerValIleGluAspLeuAlaLysAspIleThrThrPros 573
QY      1460 ACGGAGCAGACAT-----GCTCTAC 1480
Db      573 erGlyLeuAspSerLeuAlaSerArgValSerAspLysLeuAspValSerProTrpPheP 593
QY      1481 CTGACACTTCT----- 1491
Db      593 roAspThrSerValGluLysGluPheIlePheGluSerGlyLeuGlySerGlySerGlyL 613
QY      1492 -----TGCTCCACCTGCTATGGCTTACTCTCC- 1521
Db      613 yeAsnValAspValIleAspTrpProTrpSerGluThrSerLeuGluLysThrThrGluP 633
QY      1521 ----- 1521

```

```

Db      633 roLeuSerLysSerTrpSerGluGluGlnAspThrLeuLeuProThrGluSerIleGluL 653
QY      1521 ----- 1521
Db      653 ysLeuHisMetTyRPhThrGluGlnMetIleGluProSerAlaHisArgTyR GlyAspG 673
QY      1521 ----- 1521
Db      673 lyProIleTyRPhThrGluGluGluSerHisValArgSerThrIleProIlePheAlaG 693
QY      1522 -----CTGTCAGAACTCCA- 1536
Db      693 luSerAlaThrGlnProThrSerLeuIleSerSerLysHisThrSerValProAspI 713
QY      1537 -----CCTTCTTTATGGCATCAAGCATCTTCTCTCTGA 1570
Db      713 leAspSerTyRSerValThrLysAlaProPheLeuLeuAla-----T 727
QY      1571 CTGATCAAGGCACACAGATACAAATGGCCACTGACCAAG-----ACAATGCTAGTACCAG 1624
Db      727 hrIleAlaAsnThrAlaSerThrLysGluThrAspGluValAsnThrLeuLeuLysLysG 747
QY      1625 GGCTCACC----- 1632
Db      747 lyMetValGlnThrGluProSerSerProLysGlyLeuAspSerLysIleSerValAlaA 767
QY      1633 -----ATCCCCACAGTATTATTTCTGCAATCAGCC 1663
Db      767 rgProAspMetGlnProValTrpThrIleLeuProGluSerAspThrValTrpAlaArgT 787
QY      1664 AACTGGCTCTGGA-----ATTTCATCTCA- 1689
Db      787 hrSerSerLeuGlyLysLeuSerArgAspThrLeuValSerThrProGluSerAlaAspA 807
QY      1690 -----CCTGCATCTTCAGATCAGACCGCATCAAGTG 1720
Db      807 rgLeuTrpLeuLysAlaSerMetThrGlnProAlaGluLeuProProThrThrHisSerI 827
QY      1721 CAGGTGGCGAAGAT-----ATGGTCAGACACCTAGATGAATGATCTGCTGCACA 1771
Db      827 leGlnLeuGluAspGluValIleMetAlaValGlnAsnIleSerLeuGluLeuAspGlnV 847
QY      1772 CTCTCTCCCATCTGAGGTACAGAGCTCAGCGAA-----T 1807
Db      847 alGlyThrAspTyRThrGlnProGluLeuThrGlnGlnGlnAsnGlyLysValAspSerT 867
QY      1808 ATGTTTCTGCTCCAGATCATTTCTTGAGGATACCACTCTCTCTCCTCAGCTTTACGATTA 1867
Db      867 yrValGluMetProThrHisValHisTyRThrGluMetProLeuValAla---GlnProT 886
QY      1868 TCACCACTAGTTCTATGACCATTTGCCCAAGGCGGAGAGCTGGTAGTGTCTTCTCAGTC 1927
Db      886 hrLysGlyValLeuLysSerArgThrGlnThrAlaGlyAlaLeuValValPhePheSerL 906
QY      1928 TGGGTGTTCTTAACATGGCTTCTCCACAGACCTGTTCAACAGAGCTCTCTCGAGTACC 1987
Db      906 euArgValThrAsnMetLeuPheSerGluAspLeuPheAsnLysAsnSerLeuGluTyL 926
QY      1988 GAGCTCTGGACACCAATTCACACAGCTGTGTTCCATCTATCTACATCCCAATCTTACAG 2047
Db      926 ysAlaLeuGluGlnArgPheLeuGluLeuLeuValProTyRLeuGlnSerAsnLeuSerG 946
QY      2048 GATTTAAGCAACTTGAATACTTAACTTCAGAAACGGGAGTGTGATTGTGAATAGCAAAA 2107
Db      946 lyPheGlnAsnLeuGluIleLeuAsnPheArgAsnGlySerIleValValAsnSerArgV 966
QY      2108 TGAAGTTTCTAGTCTGTCGCTATTAACCTCACAGGCTGTGCACGGGCTGTGGAGG 2167
Db      966 alLysPheAlaGluSerValProProAsnValAsnAsnAlaIleTyRMetIleLeuGluA 986
QY      2168 ATTTTCTGTTCTGTCAGACCCCAACCACTCTGGAATAGACAGCTACTCTCTCAACA 2227
Db      986 spPheCyThrThrAlaTyRThrMetAsnLeuAspLysPheTyRSerLeuAspV 1006

```





Db 1042 TyrProGlyTyrLeuSerValAspGluLeuProCysGlnSerLeuCysAspLeuGlnPro 1061  
 QY 2374 GGCTCTGTGGCCCTGGCACAAGGAATCGAGAGTCTCCAGGAAAGGAGTCTCATGC 2433  
 Db 1062 AspPheCysLeuAsnAspGlyLys---CysAspIleMetProGlyHisGlyAlaIleCys 1080  
 QY 2434 AGG 2436  
 Db 1081 Arg 1081

## RESULT 15

O46616 PRELIMINARY; PRT; 185 AA.  
 AC O46616;  
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Interphotoreceptor matrix proteoglycan 150 (Fragment).  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Macaca.  
 OX NCBI\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Retina;  
 RA Kuehn M.H., Hageman G.S.;  
 RT "Characterization And Complete cDNA Sequence Of IPM 150, A Novel Human  
 RT Photoreceptor Cell-Associated Chondroitin-Sulfate Proteoglycan.";  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF047491; AAC03788.1; -;  
 DR GO; GO:0004872; F:receptor activity; IEA.  
 KW Receptor.  
 FT NON\_TER 1 185  
 FT NON\_TER 185 185  
 SQ SEQUENCE 185 AA; 21204 MW; 0214806B328C36C3 CRC64;

## Alignment Scores:

Pred. No.: 4.77e-53 Length: 185  
 Score: 841.00 Matches: 166  
 Percent Similarity: 85.44% Conservative: 10  
 Best Local Similarity: 80.58% Mismatches: 8  
 Query Match: 14.31% Indels: 22  
 DB: Gaps: 1

US-10-007-270-1 (1-3330) x O46616 (1-185)

QY 347 TTTTTCACACGGGGTAAAGTCTCTCCACAGGATCCATGAACAGATTTAGACAGT 406  
 Db 2 PhePheProAsnGlyValLysValCysProGlnGluSerMetLysGlnIleLeuAlaSer 21  
 QY 407 CTTCAAGCTTATTAGATTGAGAGTGTCTCAGGACAGCATATGGGAGCATATCGGATC 466  
 Db 22 LeuGlnAlaTyrArgLeuArgValCysGlnGluAlaValTyrGluAlaTyrArgIle 41  
 QY 467 TTCTCGATCGCATCCCTGACACAGGGGAATATCAGAGCTGGGTGAGCATCTCCACAGAG 526  
 Db 42 PheLeuAspArgIleProAspThrGlyGluTyrGlnAspTyrValSerPheCysGlnGln 61  
 QY 527 GAGACCTTCCTCTTGTGACATTCGAAACCTTCAGCAATCCAGGAGACCTGGAT 586  
 Db 62 GluThrPheCysLeuPheAspIleGlyGlnAsnPheSerAsnSerGlnGluHisLeuAsp 81  
 QY 587 CTTCTCCACAGAGAATAAACAAGAGAGTTTCCTCGACAGAAAGATGAATATCTGCA 646  
 Db 82 LeuLeuGlnArgIleLysGlnArgSerPheProGluArgLysAspGluValSerThr 101  
 QY 647 GAGACACATTGGGAGAGCTGGTGAACCATTTCTATTTCAACAGCAATCTACATTCA 706  
 Db 102 GluLysThrLeuGlyGluProSerGluThrIleValValSer--Thr----- 116  
 QY 707 AAGACTTGGGCAGTATTCTAAGAAACCTTCAGAGAGCAATTCAGATGTTGCCACG 766  
 |||||

Db 117 -----AspValAlaSerV 121  
 QY 767 TCTCACTTGGGCTTTCCCTCTCCTCCTGATGACACCTCTCTCAATGAATTTCTCGATA 826  
 Db 121 alSerLeuGlyProPheProValThrProAspAspThrLeuLeuAsnGluIleLeuAspA 141  
 QY 827 ATACACTCAACGACACCAAGATGCTTACACAGAAAGAGAAACAGAAATTCGCTGTG 886  
 Db 141 snAlaLeuAsnAspThrLysMetProThrThrGluArgGluThrGluLeuAlaValSerG 161  
 QY 887 AGGAGCAGAGGGTGGAGCTCAGCGTCTCTCTGTGTAACACAGAGTTCAAGGCAGAGTCG 946  
 Db 161 luGluGlnArgValGluLeuSerIleSerLeuIleAsnGlnArgPheLysAlaGluLeuA 181  
 QY 947 CTGACTCCCGAGTCC 960  
 Db 181 laAspSerGlnSer 185

Search completed: March 4, 2004, 18:52:41  
 Job time : 230 secs